Analysis of the Bayfield Treatment Model and Educational Outcomes

Larry S. Sanders, Ph.D.
Chairman/CEO Bayfield
sanders@bayfield.net

Robert J. Fulton, M.S.W.
PRO.FILES
Social Research and Outcomes Evaluation
r.fulton@rogers.com

January 2008

Analysis of the Bayfield Treatment Model and Educational Outcomes

1.0 Introduction:

During the past decade there has been increasing interest in ways to effectively treat aggressive adolescent males. National newspapers report the story of a13 year old boy that is tried in the adult court system for stabbing an elderly man for pocket change. This child was once in a safe treatment environment, developing relationships and subsequently discharged against clinical advice. Children are on their own trying to survive on the street by running drugs, prostituting, stealing and begging. The over-worked, under-resourced child welfare system is flooded with referrals and child protection emergencies.

A nine-year old child reports that he has been his mother's lover for the past two years and tells of how she trained him to satisfy her needs. The boy is placed in a foster home where he sexually assaults a three year old. He is swiftly removed and placed in a home with no children where he reportedly engaged in sexual behaviours with the family pet. His biological mother continues to make contact with the boy exacerbating the situation as he is moved from home to home. Fourteen placements later, the confused and rejected child has finally found his way to a residential setting where staff are trained to treat children with sexual issues.

Referrals to per diem funded agencies have doubled in the last five years resulting in waiting lists for service. The complexity of issues the child brings to treatment requires comprehensive assessment and evidence based treatment interventions.

The study takes place at Bayfield, a rural residential treatment facility established in 1973, that provides care and treatment to a population of challenging adolescent males.

The setting offers a continuum of care ranging from the most to least secure environments. The majority of the children attend Bayfield's school, with a teacher/child care therapist-student ratio of one to five, and participate in prescribed day treatment activities. An equestrian program, beef operation, woodworking shop, life skills and recreation programs may be part of each child's treatment plan.

These children have been bounced from 'pillar to post' and some have seen as many as 36 different unsuccessful placements prior to admission to Bayfield.

The setting serves children who present with significant emotional, behavioural, and psychiatric problems and who are between seven and 18 years. The average age is 12.2 years at admission. The average length of stay at Bayfield is 14.2 months, on the main campus, and 18.4 months in the community programs. Each child is admitted to Bayfield's main campus program for assessment and stabilization prior to placement in a community program. Thus, a child could reside at Bayfield for up to three years. One of the goals is to return the child to his home community, when the treatment plan has been completed.

The 180 staff complement includes psychiatry, psychology, social work, teachers and childcare workers. It is clear from the admission diagnostics that most children referred to Bayfield suffer from an attachment disorder and experience significant educational delays. The underlying assumption of attachment theory is that the experience of the infant and his early impressions of life have a dramatic impact on the kind of person he will become. This idea had its origins in the works of Plato in 400 B.C. However, it was Freud, in 1910, who cemented the idea in the field of psychology. "The very impressions which we have forgotten have nevertheless left the deepest traces in our psychic life and acted as determinants for our whole future development" (Freud as translated by Clarke & Clarke, 1976).

Although policy makers have accepted these assertions, programs for youth with attachment disorder have been under-funded and subsequently underdeveloped. During the 1980's, a great deal of research cast doubt on the validity of these assumptions that the early years predetermine future psychological development.

"Continuity between infancy and maturity undoubtedly exist, but the residual affects early experience on adult behaviour tend to be quite slight because of the maturation or changes that take place during middle and later childhood, and because of the effects of beneficial and adverse experiences during all the years after infancy" (Rutter, 1980, p. 811).

A growing number of scientists have been attracted to the life-span approach of human development. The past is only one determinant in development because the outcomes of the past are continuously transformed by new experiences, (Belsky, 1991; Chisholm, 1996). It is this experience that Bayfield intends to capture through developing meaningful relationships with the children, in order to bond with them and help them to attach. We have also learned that a key indicator of a child's success is reflected in his ability to perform in an academic environment.

The three interdependent components of Bayfield consist of residential services, clinical services and education. These components share a similar conceptual framework described in a single treatment plan. The essence of the service delivery system assists the child in developing relationships with adults, peer relationships, socialization skills, self-worth, appropriate behaviour, life skills, stress management and

stable emotional functioning, with an outcome of performing at an appropriate academic grade level.

Limitations

The study is limited by the available sample adolescent males located in one setting, with no control group for comparative purposes. Considering the type of psychopathology that some children eventually exhibit, some may not be capable of changing significantly. Some changes in the child's perception may be attributed to maturation. Finally, there are subjects who may wish to not participate in the study and others have been discharged prior to the completion of treatment limiting the sample. Different clinical staff has administered the tests and will bring their unique bias to the study. This bias may be driven by their interpretation of the desired outcome, creating false positive or false negative results. During test administration some children will 'fake good' or 'fake bad'.

The Study

This is phase two of a multi-year study of child centred outcomes related to young people receiving treatment at Bayfield. This study has been designed on discoveries of our findings of previous research and our evidence based practices. Our recent research includes, *Attachment of Adolescent Males in a Residential Treatment Setting* (Sanders, L., 2003), *Predictors of Academic Achievement for at Risk Adolescent Males, (*Sanders, L., & Jamieson, J., 2004), *Educational Achievement and Attachment – Bayfield School Outcome Study,*(Sanders,L. & Fulton,R., 2006).

Bayfield is a residential treatment centre with approximately 110 male children and youth living in 8 treatment homes and 14 next step treatment homes. Forty beds are located on the main campus. At any one time, up to 100 children attend the Bayfield School located on the main campus.

2.0 Research Objectives:

The study has the following objectives:

(1) To describe the level of academic achievement on admission, as measured by the *Wechsler Individual Achievement Test* or WIAT.

(2) To discover the pattern of change in academic skills over time.

(3) To identify what factors in the social background of the child contribute to his ability to learn.

(4) To identify what aspects of emotional disturbance (behaviour, psychiatric symptoms, etc.) affect his ability to learn.

(5) To determine the clinical outcomes, i.e., attachment, behaviour, social competence, emotional and psychopathology that are significantly related to academic advancement.

(6) To determine the outcomes of the Bayfield Treatment Model.

3.0 Measures:

The presenting problems, needs and outcomes of the children placed have been measured consistently for up to ten years across the following domains:

3.1 Educational Achievement:

Educational achievement was measured by the WIAT that contain several scales related to different aspects of the child's academic functioning. The WIAT scales include:

- Reading composite
- Weighted reading score
- Standardized score of reading (relative to a normative sample)
- Reading percentile (relative to normal grade levels)
- Mathematics – raw score
- Mathematics standard score
- Mathematics percentile
- Oral learning raw score
- Oral learning standard score
- Oral learning percentile
- Supplemental reading score
- Supplemental "two word" reading
- Reading speed
- Reading grade level (standardized on Canadian norms)
- Math grade level (standardized on Canadian norms)
- Oral Learning grade level (standardized on Canadian norms)

3.2 Clinical Assessment:

Children are assessed by a wide range of clinical instruments during treatment. However, four clinical instruments are administered to all children:

(1) the *Conners' Global Index* (CGI), a measure of impulsivity, hyperactivity and emotional lability

(2) the *Children's Global Assessment Scale* (CGAS), the children's version of axis V of the DSM-IV, that measures how well the child functions within the major roles of his/her life

(3) the *Symptom Assessment Checklist* (SA-45), provides a measure of psychiatric symptomatology for adolescents

(4) the *Feelings, Attitudes and Behaviour Checklist* (FAB-C), measures the inner distress of children from age 8 to age 13 years.

(5) The *Level of Assistance Scale* (LOAsocial),which measures the amount of assistance (or adult mediators) required to enable the young person to function in terms of the daily activities of living. A child who requires one-to-one supervision to get dressed for example would have the highest score (4) in the sub-domain, *grooming and dressing*. A child who needs no assistance whatsoever would have a score of zero.

3.3 Personal Adversity and Distress

There are two instruments used to measure personal adversity in the background of the children; the history of physical and sexual abuse, and current stress levels.

(1) The *Objective Stressors Checklist* (OS), a measure of the number of daily hassles, worries about the macros social system, health fears, life altering events and serious trauma that the child can identify

(2) The *Sociodemographic Checklist* (SD), a list of 15 serious stressors apparent in the clinical record of the young person that have been shown in longitudinal studies to increase the risk of social dysfunction in young adulthood.

3.4 Attachment

Attachment is measured by the *Parental Bonding Instrument* (PBI) that measures the adolescent's self appraisal of how much he was cared about and controlled during his childhood. This instrument has two scales:

(1) The *caring scale* measures how much the young person feels he was cared about by the "closest parent figure" in his life.

(2) The *overprotection scale* measures how much the young person feels he was treated fairly by his closest parent figure.

Although the actual life events of a person's early childhood does not change, the teenagers' self appraisal of their childhood is subject to change in treatment. Abusive parents may not change significantly while the child is in treatment but the child may adopt a new perspective of his parents, characterised by such changes as greater understanding, tolerance and acceptance of his parents. These changes are measured by the PBI.

4.0 Participants

This study began in the spring of 2005. As of October 2007, 146 young people attending classes have been assessed on the WIAT test of educational achievement. Eight months later (244 days) on average, 117 children were tested a second time on the same instrument and after another similar waiting period, 66 children were tested for a third time. Recently 18 young people have been tested for fourth time since their treatment began. When the first academic tests were completed, the children had spent 1.4 years in treatment and 145 of the children assessed by the WIAT were also assessed using measures of emotional disturbance. The program of outcome measurement had begun at Bayfield for the residential treatment service seven years before this research project started. The children included in this study are part of the larger residential treatment study that has gathered data on 252 children.

5.0 Research Design

The dataset for this project consists of multiple waves of test scores. There are up to 10 waves of clinical and attachment scores; there are up to four waves of educational achievement scores. The dataset also includes descriptive information about the children, including: (1) the grade in which they are enrolled, (2) their age at each test period, (3) whether or not the child was diagnosed as learning disabled and (3) specific adversity in their background, e.g. history of abuse, age when abuse started, parental history of criminality, mental illness, mental retardation and substance abuse.

The relationship between all of the variables was explored, including correlation and factor analysis. In addition, basic descriptive statistics was obtained for all variables.

The difference between scores at each subsequent wave of testing after the initial test was completed and *the effect size* was computed. The formula for the effect size of pre-post comparisons for matched pairs is as follows: Effect Size (d) = (mean2 – mean1) / standard deviation of T1; an effect size of .2 is small, .5 is medium and .8 or larger is large. The effect size creates a standardized way of presenting the changes over time.

The entire dataset is a time series, meaning that the order of test results is apparent and the time between testing can be computed exactly. This provides a method to observe changes in earlier periods on the variables to predict changes in clinical measures and educational achievement. In particular, we are interested in whether or not improvement in attachment scores predicts improvement in achievement as intervention theory suggests. We have used multifactor regression analysis to explore these possible relationships.

6.0 The Baseline Condition of Children Admitted

6.1 Descriptive Data on the Children

TABLE 1: ACADEMIC BASELINE SCORES

	N	Minimum	Maximum	Mean	Std. Deviation	
ClientAge	145	7.4	17.7	13.2	2.1	age when first tested on academic achievement
ReadComp	141	-	192.0	28.0	18.4	Reading composite
ReadWieght	141	14.0	213.0	116.8	35.1	Weighted reading score
ReadStandard	141	40.0	125.0	76.8	21.5	Standardized score of reading
ReadPerc	141	-	95.0	19.5	25.7	Reading percentile (relative to normal grade levels)
MathRaw	144	7.0	64.0	39.6	10.6	Mathematics – raw score
MathStandard	144	40.0	130.0	71.9	19.6	Mathematics standard score
MathPerc	144	-	98.0	12.6	21.4	Mathematics percentile
OralLraw	143	8.0	39.0	25.8	6.0	Oral learning raw score
OralLStandard	143	40.0	136.0	83.2	18.8	Oral learning standard score
OralLperc	143	-	99.0	23.7	26.0	Oral learning percentile
SupRead	98	1.0	4.0	1.8	1.0	Supplemental reading score
SupTwords	91	1.0	4.0	2.0	1.2	Supplemental "two word" reading
SupReadSpeed	89	1.0	4.0	2.2	1.2	Reading speed
ReadGradeD	141	-	13.0	4.9	3.1	Reading grade level (standardized on Canadian norms)
OralGradeD	143	-	13.0	5.0	3.4	Oral Learning grade level (standardized on Canadian norms)
MathGradeD	144	-	12.9	4.9	2.5	Math grade level (standardized on Canadian norms)

6.1-a Interpretation of Academic Scores

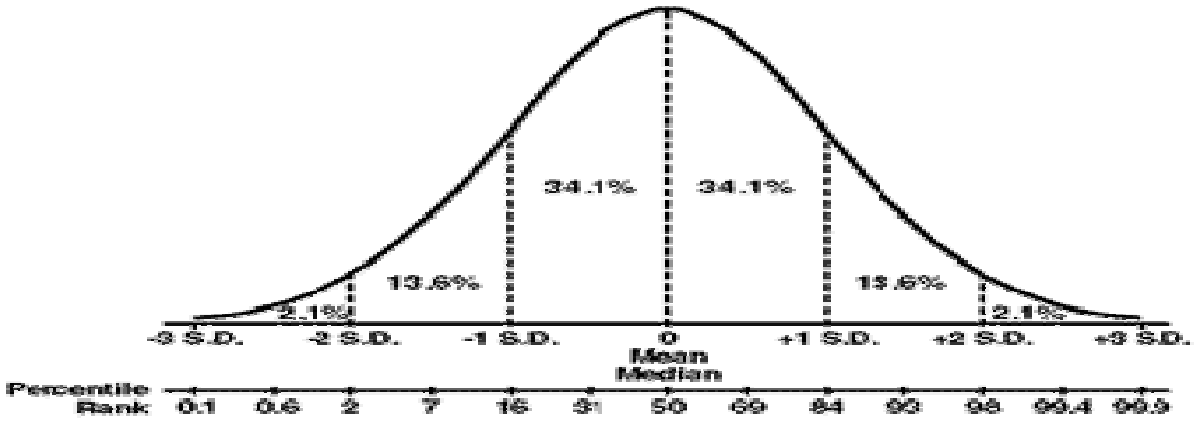
The table above reveals the following findings:

(1) The average age of the children tested is 13.2 years and under normal circumstances they should be achieving at a grade seven level in all subjects.

(2) In reading the children have a standard score of 76.8 (compared to the normative mean score of 100). This means that the children in Bayfield have a reading ability below the 20th percentile of a typical classroom for their age group in Ontario.

(3) A note of caution, however, is that the standard deviation specific to the Bayfield population is quite large (21.5). This means that at least 16% of the children attending Bayfield show reading abilities at or above the average child.

Below is a graph of the normal curve. Note the "normative" mean for the WIAT standard scores in the above table is 100[1].



(4) The standard score in math is lower (71.9). This places the Bayfield population below the 13th percentile in math skills.

(5) Conversely, the children in Bayfield show slightly better oral learning skills, with a standard score of 83.2, or just below the 25th percentile.

[1] This is the same scale as the IQ test, which set the normative mean at 100. The standard scores for most clinical assessment tools are based on a different scale, with a normative mean of 50.

(6) Overall, the children are functioning at a *grade 4.9 level* in math and reading and at grade 5.0 in oral learning. As noted above, the children should be within a grade seven level. This means the children in Bayfield are 2.1 years behind their peers in academic skills on admission.

6.2 Generalizability of Findings

One of the key research objectives is to explore the relationship between changes in the child's clinical and social status – especially attachment – and his ability to make progress academically. It would be useful if we could accept that these relationships apply to all children placed in Bayfield, not just the children served over the past two years.

This section of the report considers the question: *is there any difference in family background, clinical test score and attachment between the children who were part of the Study and those who were placed earlier and were not tested on the WIAT*?

An analysis of variance (ANOVA) was performed on 23 clinical and social variables, comparing children who are were part of the Study and those were served before the study began. The two groups were compared on the following variables:

TABLE 2: THE GROUP BEFORE WIAT TESTING BEGAN COMPARED TO THE WIAT GROUP

		sig diff	subgroup	subgroup
			before the study	in the study
Clinical Scales				
Conners' Global Index	CGI			
Children's Global Assessment Scale	CGAS	*yes*	49.48	46.14
Psychiatric symptoms	SA45			
Feelings, attitudes and behavior	FabC			
LOAsocial	LOAsocial	*yes*	19.39	24.97
history of adversity and personal distress				
Objective Stressors Checklist	OS			
Socio Demographic Checklist	SD			
see details below				
years of poverty & hardship	*hardship*			
history of sexual abuse	*sexual abuse*			
history of physical abuse	*physical abuse*			
suicide of close family member	*suicide*			
jail of close family member	*jail*			
mental illness of close family member	*psych*			

parent has mental retardation	*MR*			
substance abuse of close family member	*fam drug*	*yes*	70% positive	53% positive
history of rape of close family member	*rape*			
history of drug abuse by the child	*drugs*			
child has learning disability	*lrn dis*			
parent of child was a teen mom	*teen mom*			

Attachment Scales	
Care T	Care T
OverP T	OverP T

Other categorical variables, such as aboriginal identity and minority status were checked using a chi-square analysis and no significant differences emerged. In summary, there were only three variables which showed a significant F-ratio: CGAS, LOAsocial and a history of drug abuse by a close family member. The children in the study have slightly greater functional impairment and more need for adult assistance in daily living compared the children served earlier. The children in the study have less drug abuse in their family background. These are not materially significant differences and we conclude the findings can be generalized to all children placed in Bayfield over the past seven years.

6.3 Relationship between Clinical and Educational Variables

The treatment and educational staff Bayfield believe that:

> *by promoting positive changes to the child's attachment patterns, personal distress, psychiatric and behavioural symptoms and social adjustment the child will be make significant progress in academic skills – since he will have more psychological energy to invest in the classroom and greater trust in and attentiveness to the teacher*

The first step in verifying this treatment framework is to determine if these major domains are independent, i.e., not correlated on admission. The independence demonstrates that we are measuring different variables. Secondly, the independence sets up the experiment; testing the assumption above requires us to look at the child at a time before the "supposed effect" occurred. If the factors listed above are already highly associated with academic skills when children are tested for the first time, then the "outcome" has already occurred and there is little room left to run an experiment.

All of the test scores on children at T1 were processed through a confirmatory factor analysis; four distinct and independent domains emerged as illustrated through the following table:

TABLE 3: FACTOR ANALYSIS OF BASELINE SCORES

academic scale		clinical scale		attachment scale		distress scale	
Rotated Component Matrix(a)							
	1		2		3		4
ReadGradeD	0.963	SA45	0.691	OverP T	0.897	OS	0.791
ReadStandard	0.942	CGAS	-0.689	Care T	-0.770	SD	0.705
ReadWeight	0.939	CGI	0.646				
OralGradeD	0.880	LOAsocial	0.607				
MathStandard	0.875						
MathGradeD	0.874						
OralStandard	0.866						
ReadComp	0.838						

The numbers in the cells above show the correlation between the individual scales, e.g., SA 45, CGAS, CGI and the common underlying factor, i.e., the "clinical scale". The correlation between the individual scale and the common underlying factor are much higher than the correlation between the different measures. The common factors, academic scale, clinical scale, attachment scale and distress scale are not correlated with each other at T1. They are *independent* and *distinct* variables.

This is the first step in validating the treatment framework discussed above. The theory of change at Bayfield presumes these underlying factors will become correlated between each other and dependencies will become apparent. In particular, we are testing the hypothesis that later academic growth is dependent on prior changes to the attachment scales.

6.4 Social (or family) Variables related to Academic Functioning on Admission

Certain aspects of the children's background and identity are strongly related to their academic abilities on admission: either making their academic performance worse or better. These findings were discovered by analysis of variance between the child's background and each one of the academic measures. Note there are very few aspects of the background of the children that predict their academic ability.

6.4-a Children with Aboriginal Identity

TABLE 4: ANOVA TABLE COMPARING NATIVE AND OTHER CHILDREN

		N	Mean	F-ratio	Sig
ReadComp	others	52	29.6	7.037	0.010
	Aboriginal	15	21.0		
	Total	67	27.7		
ReadWeight	others	52	127.4		

		N	Mean	F-ratio	Sig	
				12.275	0.001	
	Aboriginal	15	93.1			
	Total	67	119.7			
ReadStandard	others	52	81.5	5.707	0.020	
	Aboriginal	15	68.5			
	Total	67	78.6			
ReadPerc	others	52	21.9	4.419	0.039	
	Aboriginal	15	8.0			
	Total	67	18.8			
MathRaw	others	52	43.2	18.183	0.000	
	Aboriginal	15	30.8			
	Total	67	40.4			
MathStandard	others	52	76.3	6.202	0.015	
	Aboriginal	15	63.8			
	Total	67	73.5			
MathPerc	others	52	14.4	3.501	0.066	
	Aboriginal	15	4.7			
	Total	67	12.2			
OralLraw	others	52	28.0	20.680	0.000	
	Aboriginal	15	20.3			
	Total	67	26.3			
OralLStandard	others	52	88.9	11.602	0.001	
	Aboriginal	15	71.3			
	Total	67	85.0			
OralLperc	others	52	30.8	8.232	0.006	
	Aboriginal	15	10.3			
	Total	67	26.2			
SupRead	others	41	1.9	1.253	0.269	no relationship
	Aboriginal	9	1.4			
	Total	50	1.8			
SupTwords	others	37	1.9	0.810	0.373	no relationship
	Aboriginal	8	2.3			
	Total	45	1.9			
SupReadSpeed	others	37	2.1	0.967	0.331	no relationship
	Aboriginal	8	2.6			
	Total	45	2.2			
ReadGradeD	others	52	5.8	12.261	0.001	
	Aboriginal	15	2.9			

		N	Mean	F-ratio	Sig	
	Total	67	5.2			
OralGradeD	others	52	6.2	17.914	0.000	
	Aboriginal	15	2.3			
	Total	67	5.4			
MathGradeD	others	52	5.7	13.720	0.000	
	Aboriginal	15	3.1			
	Total	67	5.1			
CGI	others	85	64.7	0.772	0.381	no relationship
	Aboriginal	30	62.1			
	Total	115	64.0			
CGAS	others	85	47.5	0.050	0.823	no relationship
	Aboriginal	30	48.1			
	Total	115	47.7			
SA45	others	50	59.4	1.743	0.192	no relationship
	Aboriginal	11	55.5			
	Total	61	58.7			
OS	others	80	3.3	0.002	0.963	no relationship
	Aboriginal	28	3.4			
	Total	108	3.3			
SD	others	85	3.8	2.731	0.101	no relationship
	Aboriginal	30	4.4			
	Total	115	3.9			
FabC	others	32	54.9	2.009	0.163	no relationship
	Aboriginal	15	59.4			
	Total	47	56.3			
Care T	others	53	52.6	0.559	0.458	no relationship
	Aboriginal	10	50.2			
	Total	63	52.2			
OverP T	others	53	52.5	0.037	0.849	no relationship
	Aboriginal	10	51.8			
	Total	63	52.3			

In summary, 13 out of 16 academic measures are significantly worse for children of Aboriginal identity than all others. There are no differences, however, between children of Aboriginal identity and other ethnic groups on clinical variables, attachment variables or indicators on adversity and distress. A separate examination of each unique

ethnic group did not produce a similar result. Aboriginal children have a special difficulty succeeding in the Anglo education system (Powers, K., Potthoff, S. J., Bearinger, L. H. & Resnick, M.D. 2003).

Powers et al (2003) state in their literature review as to the underlying causes of failure in school by Aboriginal students, that "A frequently cited deleterious effect of cultural discontinuity is the challenge it poses to students' opportunities to form attachments with school personnel. Establishing trusting relationships with teachers and other school staff is an asset to school children. Masten (1994) observed that "schools ... not only provide knowledge and teach problem-solving skills; they provide a setting where children can become connected with caring, competent adults" (p. 15). Results from the National Longitudinal Study of Adolescent Health indicate that perceived teacher caring and fairness are critical components of "school connectedness" (Resnick et al., 1997). Such connectedness is significantly associated with better emotional health among students as well as lower levels of involvement in health-risking behaviours such as substance use and self-directed and interpersonal violence (Resnick et al., 1997). Furthermore, Pavel and Padilla (1993) found students' sense of social integration to promote postsecondary American Indian/Alaska Native students' school persistence. American Indian/Alaska Native students, like all other students, learn better in a supportive social context than from impersonal teachers (Murdoch, 1988)." (Powers, 2003, p. 18)

The causal hypotheses cited in the literature to explain Aboriginal failure in school has been generalized to apply to all children in Bayfield.

6.4-b Diagnosis of Learning Disability

Noted in Section 6.1, the children are on average 2.1 plus years behind their peers in academic attainment when first admitted to Bayfield.

TABLE 5: CROSS TAB ABORIGINAL AND LEARNING DISABILITY

		Aboriginal		**Total**	
		no	*yes*		
Lrn dis	*yes*	55	21	76	77%
	no	23	7	30	23%
Total		78	28	106	

		Aboriginal		
Lrn dis		*no*	*yes*	
by col	*yes*	71%	75%	of aboriginal are LD
	no	29%	25%	

by row	*yes*	72%	28%	of LD are aboriginal
	no	77%	23%	

This table means that 77% of the children in Bayfield are admitted with a diagnosis of learning disability provided by the public school system in which they were placed previously. Moreover, 28% of the children with LD are aboriginal. The chi-square statistic is not significant meaning Aboriginal children have the same chance of being diagnosed with LD as anyone else.

Comparing the standard scores for the two groups, it is clear that the Aboriginal children have lower standard scores than the total group of Learning Disabled children, specifically:

		N	Mean
ReadStandard	Aboriginal	15	68.5
	Learning Disability	70	71.7
MathStandard	Aboriginal	15	63.8
	Learning Disability	70	67.7
OralLStandard	Aboriginal	15	71.3
	Learning Disability	70	79.4

Children with diagnosed learning disability show a significant F-ratio and lower educational attainment compared with children not diagnosed. In conclusion, there are two social conditions, existing prior to admission to Bayfield, which depress the scores on most scales on the WIAT[2].

6.4-c The Impact of Other Social Factors on Academic Attainment

Other adverse social conditions affect individual academic and clinical scales. In some cases, an adverse social factor affects academic attainment in a counter-intuitive direction; the children appear to have an academic advantage. The following table describes the impact of pre-existing social conditions.

All children admitted to Bayfield have at least one condition of adversity on this list; some have nine of these factors in the background and the Bayfield average is 4.0.

[2] The exceptions are the supplementary tests for reading speed, reading two words and reading in general, which are not significantly different when either LD or Aboriginal are entered as the grouping factor. In the case of the clinical scales, only the CGI is significantly worse for the LD children; whereas there is no difference on this variable when Aboriginal identity is the grouping factor.

Due to the prevalence of severe environmental distress, the entire population of Bayfield experiences a baseline of problems and symptoms in the educational, clinical, personal distress and attachment domains. The table describes how each single stressor affects the baseline amplifying or suppressing it to some degree.

TABLE 6: THE EFFECT OF EACH INDIVIDUAL ENVIRONMENTAL STRESSOR ON THE CHILDREN

	Percent yes	*academics affected*	*objective stress*	*clinical effect*
hardship	32.8%	no *added* effect	no added effect	no added effect
sexual abuse	38.6%	no added effect	**4.4 vs 2.5**	no added effect
Physical abuse	64.6%	no added effect	**4.0 vs 2.5**	no added effect
suicide of close family member	2.6%	no added effect	no added effect	no added effect
jail of close family member	27.0%	math standard score is 10 points higher and grade level is advanced by 1 + years	**4.6 vs 3.0**	no added effect
psychiatric hospitalization of close family member	9.0%	no added effect	no added effect	no added effect
close family member is MR	12.7%	oral standard score is 12 points lower	no effect	FAB-C (9 points higher) in the borderline clinical range
close family member with substance abuse disorder	61.4%	no added effect	no added effect	
rape of close family member	24.3%	read standard score is 10 points higher and grade level is advanced by 1.5 years	**4.3 vs 3.2**	slightly more functional (CGAS) by 3 points
child abuses drugs	12.7%	oral and math standard scores are higher by 15 points and reading, oral and math grade levels are 2-3 grades higher	no added effect	no added effect
child is learning disabled	77.2%	significant effect	no added effect	slightly more hyperactive (CGI) by 5 points
parent was teen mom	23.8%	no added effect	no added effect	no added effect

In general, these pre-existing conditions had affected the academic, clinical and attachment functioning of the children in Bayfield long before being admitted to Bayfield. The most striking noticeable differences that certain individual stressors have on children are as follows:

(a) Children who are physically or sexually abused have many more daily hassles and "objective stressors" which they identify in the "here and now" of the program.

(b) Children who have a family member that was in jail or was raped also report more current distress (Objective Stressors).

(c) Three factors in the background (family member in jail, family member raped and child has a history of abusing drugs) are associated with higher academic functioning that is close to the average functioning of their peers in society.

The children admitted to Bayfield present significant educational deficits after spending, on average, seven years in the public school system. The explanatory factors include learning disability, Aboriginal identity, severe environmental stressors including years of poverty, child abuse, and parental substance abuse. Given the initial deficit, it is a difficult challenge to help these children succeed in school and graduate successfully. The key research question concerns the power of clinical changes achieved in the residence to liberate the child so he is able to take advantage of instruction and catch up with his peers academically.

7.0 Analysis of Change

An SPSS dataset has been created that contains four waves of educational testing approximately 9 months apart. There are 117 "matched pairs" or the same child tested in wave one and wave two. The number of matched pairs, T1 to T3 or T1 to T4, decreases with each subsequent wave of testing.

Out of the initial 117 matched pairs, 70 young people were tested less than one month earlier on clinical, attachment and stress scales. An additional 30 children had been tested years earlier on at least three or more occasions on the clinical, attachment and stress scales. Finally 17 children had been tested only twice before on the clinical … scales prior to the first wave of educational testing. For these 47 children, we created a variable of the amount of change in the caring and overprotection scales over the prior history of testing. This represents changes in the key explanatory variable that pre-dates any changes in the educational attainment scores.

The final dataset includes the results of clinical testing (CGI, CGAS, SA-45, FAB-C and LOASocial), attachment scales and stress scales that occurred either within one month prior to the educational testing (70 children) or several months prior to educational testing (47 children). The dataset also includes four waves of clinical, attachment and stress scales starting from one month after admission of the child.

7.1 Methodology for Considering the Effect of Treatment

There are three ways to represent differences in test scores over time:

(a) An arithmetic difference (score at T2 minus score at T1)

(b) The t-score of the difference in means

(c) The effect size

The arithmetic difference is often presented as a percentage change using the formula: *(T2 (score) – T1 (score)/T1 (score) times 100*. The average of the individual difference scores reflects the group difference. The advantage of this approach is that the material or clinical significance of the change over time is readily apparent. The disadvantage of this approach is that it doesn't take the range of scores into account; some children may actually deteriorate while others are "super performers" and a few are "middle of the road performers". When averaged together, the "group difference" appears to be "middle of the road performers, which reflects the minority of cases.

The t-score of the difference in means[3] is provided by the formula:

$$(mean_A - mean_B)/SQRT(SEM_A^2 + SEM_B^2)$$

The advantage of the t-score in the difference of means is that it rules out "random chance" as an explanation for the arithmetic difference; this is referred as "statistical significance". The disadvantage of this approach is the difference may be statistically significance but the difference is so small that it is *clinically insignificant*.

The effect size corrects the disadvantages in the prior methods. It provides a clearer picture of the material or clinical significance of changes over time and completely equalizes the variance of change scores across the entire cohort. The effect size is given by the formula:

(T2 – T1)/SD of T1

The effect size is *change scores over time* converted to a standardized value (z-score). Because it is a standardized value, the effect scores of all tests and even rates of target behaviours can be pooled. Within an agency, the effect scores for all tests or indicators of change can be averaged for an agency-wide "change effect". In fact, effect sizes can be manipulated in simple linear ways to produce "common sense" insights. For example, if the effect size of the Bayfield classroom on mathematics standard scores is .6 in contrast with an effect size of .5 for reading, then we can conclude that Bayfield

---

[3] The term, $mean_A$ refers to the average value of the scores at T1. The term, SEM, is the standard error of the mean and is given by SD_A^2/n_A. The SEM for each group is squared, added together and the square root of this product is taken. The result is the "pooled standard error of the mean".

school is 20% more effective teaching math versus reading. Also, the effect size has its own standard deviation and median across a group of subjects. The balance of statistical applications can be validly manipulated to compare treatment effects across groups and across measures. The rule for interpreting effect sizes is provided by Cohen (1969):

> *An effect size of .2 is considered a small effect; an effect size of .5 is a medium effect and an effect size of .8 or larger is a large effect.*

Between agencies, the effect sizes can be pooled for a system wide effect size. This is very useful to pool the results of several studies using the exact same measure. As a result, the problem of low samples in any one study can be completely overcome. This is now the standard approach in meta-analysis. The first large mega-analysis using effect size was a review of the efficacy of psychological, educational and behavioural treatment by Lipsey and Wilson (1993).

There are a number of limitations to the effect size; it may be biased. Quality assurance in research practice within an agency requires one to monitor "effect size bias" before jumping to conclusions about the efficacy or lack thereof of any particular program or intervention. Lipsey & Wilson (1993) reviewed the various types of bias and its impact on "effect sizes": (a) *selection bias:* clients whose pre-treatment status is better than controls inflates the effect size because less disturbed clients generally do better; (b) *design bias*: single group pre-post designs report effect sizes that are 65% higher than comparison group designs; (c) *publication bias:* published studies report average effect sizes of .53 compared to unpublished studies of .39; (d) *small sample bias:* studies with 25 treatment and control subjects averages effect sizes of .58 and large samples (100 or more) had an average effect size of .35 for all types of treatment.

The range of interventions reviewed by Lipsey & Wilson (1993) provides a baseline for comparing outcomes in Bayfield. Additional baseline data that acts as a quasi comparison group for Bayfield is provided by Andrews & Dowden (2000) and Lipsey (1992). Lipsey & Wilson (1993) found that

(a) Behavioural self-management, social skills trainings, cognitive behavioural therapy and biofeedback/.relaxation training with problem children; all relevant clinical outcomes: grand average effect size = .61 with 43 studies

(b) Psychotherapy with children and adolescents; all outcomes: grand average effect size = .79 with 108 studies

(c) Treatment programs for juvenile delinquents: delinquency outcomes: grand average effect size = .17 with 397 studies

(d) Treatment programs for adjudicated delinquents in residential/institutional settings; all outcomes: grand average effect size = .37 with 111 studies

(e) Behavioural treatment approaches for juvenile delinquents; long-term outcomes: grand average effect size = .40 with 25 studies

Lipsey & Wilson (1993) found that different treatments worked quite well (moderate to strong effects), but this applied to heterogeneous client groups. When the effect size was calculated on various treatments for juvenile delinquents, the effect size was much smaller (small to moderate range).

Andrews & Dowden (2000) conducted a meta-analysis of correctional treatment for violent offenders; they uncovered rules that enhanced treatment effects for children and adults with severe behavioural problems. They found that:

(a) Consider variability exists in effects sizes ranging from -.22 to .68

(b) "the level of treatment services must be appropriately matched to the risk level of the offender", i.e., the *risk principle,* adherence to the risk principle causes an effect size to be .09 compared to .04

(c) The provision of service, i.e., the *human service* principle, as opposed to simply imposing criminal sanctions raises effect size from -.01 to .12

(d) When the program content targeted the known risk factors, i.e., the *criminological need* principle of violence, violence related outcomes improved from .00 to .20

(e) When behavioural treatment techniques are used ,i.e., the *general responsivity* principle, violence related outcomes improved from .1 to .19

(f) When different treatment studies were classified by experts on a scale from inappropriate service to most promising service, the resulting effect sizes were significantly different:

a. inappropriate services had an effect size of -.01 with 23 studies

b. weak services showed an effect size of .07 with 11 studies

c. promising service showed an effect size of .05 with 5 studies

d. most promising services showed an effect size of .20 with 13 studies

The meta-analytic studies of treatment effects show considerable variation; as a rule, psychotherapy and cognitive behavioural therapy show moderate to large effects with various populations; when the target population are children and teenagers with

juvenile delinquency and violence, the effects sizes tend to range from small to moderate even when the most promising treatments are tested.

7.2 Academic Effects

Academic effects were measured by examining the changes in the standard scores for reading, oral learning and math. These scores are standardized against age-related norms that vary every three months. In a normative population, the raw scores for reading, oral learning and math improve continuously as long as they receive classroom instruction. If the children improve at a "normal rate", their standard scores remain stable, reflecting the fact their "reading levels" are the same relative to their peers. If a child was functioning academically at the absolute average level throughout his entire career in school, his standard score would remain unchanged at 100 for 12 years. His raw scores would have increased steadily, as well as his grade levels. The academic effect size for this totally average Canadian child would be exactly zero throughout is entire academic career.

Therefore, an effect size of greater than zero on standard scores implies the child is learning at an accelerated pace relative to a normative sample of Canadians. If a child is behind his peers at T1, a positive academic effect would mean that he is "catching up" to his peers. Table #8 shows that the total group of children in Bayfield School are making accelerated progress.

TABLE 8: ACADEMIC EFFECT SIZE ALL CHILDREN IN BAYFIELD

						mean standard scores		
	N	Min effect	Max effect	Mean effect	Std. Deviation	time 1	time 2	Diff_t/sig
reading effect at T2	113	- 1.876	1.876	0.147	0.717	77.16	80.56	- 2.185 sig = .03
oral effect at T2	114	- 1.178	2.731	0.247	0.617	82.95	87.56	- 4.277 sig = .0000
math effect at T2	115	- 1.853	1.471	0.077	0.534	71.47	72.89	Not sig
academic effect at T2	115	- 1.240	1.521	0.161	0.432			
						time 1	time 3	
reading effect at T3	64	- 1.970	1.970	0.241	0.829	73.47	78.61	- 2.327 sig = .023
oral effect at T3	64	- 1.071	1.874	0.448	0.602	79.14	87.52	- 5.962 sig = .0000
math effect at T3	64	- 1.526	2.507	0.055	0.655	67.31	68.33	Not sig
academic effect at T3	65	- 0.604	1.675	0.246	0.489			

7.2-a Interpretation of Academic effects

The columns identified by the word "effect" are statistics related to the relevant effect score for reading … etc. The three columns on the far right apply to the standard scores only. The differences between these standard scores were also standardized into a statistic referred to as the "effect score".

At T2, there is a small effect (.2347) in changes on the oral standard score. This means that the children are learning at a faster pace than their peers in the normative sample and they are catching up to the norms in society. Conversely, there is no effect for reading and math at T2. This means that the children are learning new material in reading and math but they are just keeping pace with their peers in society. However, just *keeping pace*, in effect means they have reversed a multi-year negative trend. The actual standard scores in the far right hand column show the slow upward progress; the t-score in the difference shows that the tiny improvement in reading is still statistically significant, even if it is not materially significant.

At T3, the number of children (n) is slightly over half of the cohort with two waves of testing; as a result the score at T1 is slightly different for the 3-wave group. By T3, there is a small reading effect (0.241) and a medium oral learning effect (.448). The difference in standard scores for reading and oral learning are statically different as well meaning that there is an almost zero chance that random chance variations are driving the difference. The overall academic effect is simply the average of the reading, oral and math effects. The overall academic effect is in the small range at T3.

The reader will also note a very large standard deviation of effect scores, which are two or three times higher than the mean effect score. This means that the academic effect (globally and in reading, oral learning and math) varies significantly child by child. Clearly, the research must account for this variation. It is possible that variation in some other measurement variable in the dataset explains the variation in academic scores. It is more likely that some hidden factor, a latent variable, is responsible. The statistical methodology in a later section of this report will present a model that explains the academic variation in terms of variables we measure (emotional and psychiatric symptoms, attachment, functional ability, etc.) and some deeper underlying factor(s). This methodology is referred to as structural equation modelling.

7.3 Clinical Effects

This table in similar in structure to Table 8; the effect scores on the left and the standard scores are in the right three columns. As above, the T1 scores vary according to whether you are looking at the 2-wave set or the 3 wave set.

TABLE 9: CLINICAL EFFECT SIZE ALL CHILDREN SERVED

						mean standard scores		
	N	Min effect	Max effect	Mean effect	Std. Deviation	time 1	time 2	Diff_t/ sig
Conners' Global Index at T2	264	- 2.480	3.137	0.329	1.124	65.46	69.97	- 4.761
Children's Global Assessment T2	262	- 3.366	2.712	- 0.376	1.072	47.53	51.56	- 5.682
Level of Assistance at T2	188	- 3.360	2.387	0.006	1.018	22.94	23.01	not sig
SA-45 at T2	119	- 2.277	1.782	- 0.285	0.817	58.86	55.98	3.800
FAB-C at T2	82	- 2.687	2.886	- 0.194	1.113	58.27	56.32	1.580
Overall clinical effect at T2	264	- 1.746	1.974	- 0.074	0.671			
						time 1	time 3	
Conners' Global Index at T3	196	- 3.501	3.282	0.362	1.232	65.21	70.17	- 4.116
Children's Global Assessment T3	195	- 3.834	3.927	- 0.490	1.185	47.04	52.28	- 5.771
Level of Assistance at T3	135	- 3.625	3.271	- 0.020	1.193	23.30	23.07	not sig
SA-45 at T3	85	- 3.070	2.079	- 0.340	0.905	58.86	55.42	3.467
FAB-C at T3	37	- 2.687	3.384	- 0.328	1.308	60.54	57.24	-1.968 sig .051
Overall clinical effect at T3	196	- 2.331	2.060	- 0.099	0.836			

The overall clinical effect is the average of the five effect sizes related to each of five clinical tests. Each one of the clinical tests measure symptoms of various types. For four of the tests, the standard numbers are expected to go down over time, showing fewer symptoms; in the case of the Children's Global Assessment Scale, progress is indicated by higher numbers on the scale. Technically speaking, the CGAS and the Level of Assistance scales are not in standard scores; they are raw numbers and the actual tests are "criterion reference tests" rather than normative tests. The variation in scoring approaches for each test shows the value of the statistic "effect size"; it creates a level playing field for evaluating differences on all manner of tests and numbering systems.

7.3-a Interpretation of Clinical Effects

The overall clinical effect size is zero; however, this is the end result of forces pulling in opposite directions rather than a uniform case of no changes in symptoms. If you look at the right hand columns, the story is very clear:

(a) Children become more hyperactive and inattentive over time as measured by the Conners' Global Index; the mean effect size is "positive" meaning an increase in symptoms.

(b) Children become more functional; they fit into society better as measured by the Children's Global Assessment Scale and the effect size is in the moderate range. Functional adaptation is the best predictor of future success (Rutter & Sandberg, 1985).

(c) The level of assistance scale is not changing at T2 or T3; this scale measures how much adult support is required to assist the child in the social dimensions daily living. The standard deviation of the LOASocial scale is quite large which indicates that some groups of children are improving and other groups are not. This is a logical outcome; if the adult assistance required is due to brain injury, e.g., Fetal Alcohol Spectrum Disorder, then we cannot expect much change in this scale.

(d) The SA-45, a self report instrument, shows a small improvement at both T2 and T3 that strengthens over time. This improvement is statistically significant as well. However, the test for the 2-wave cohort was administered to 119 teenagers; across the larger population (almost every resident in Bayfield over 13 years of age), the baseline score is within the average range of the normative population (t-score of 58.86) well below the clinical range of 70+; the baseline score is close to one standard deviation above the normative mean. This means that the teenagers were reporting a fair number of symptoms, i.e., anger, depression, self doubts, hostility, but these symptoms were not as numerous as one would see in a population of teens in a psychiatric hospital. Nevertheless, these teenagers did make substantial improvement, but the effect size is still small because there is *not much room for improvement* in a population of teenagers who are already within the average range relative to a normative sample.

(e) The FAB-C (*feelings*, *attitudes* and *behaviour),* a self report instrument, shows a small improvement at both T2 and T3; almost every child at Bayfield who was too young for the SA-45 was administered the FAB-C instead; no resident was administered both the FAB-C and the SA-45. The FAB-C effect sizes show an even greater standard deviation than the SA 45. This means that some children are improving much more than others groups. The baseline score for the entire group of children being administered the FAB-C (n = 82) is 58.27 within one standard deviation of the mean score for the normative sample. There is not much room for improvement when the respondent is this close to the normative mean; the small effect size is to be expected and

reflects a positive change in feelings and attitudes overall, which is a good clinical indicator.

7.4 Effect Size in the Measure of Attachment

TABLE 10: BONDING EFFECT SIZE

						mean standard scores		
	N	Min effect	Max effect	Mean effect	Std. Deviation	time 1	time 2	Diff_t/sig
caring effect at T2	134	- 3.788	4.477	- 0.092	1.335	51.53	50.05	not sig
over -controlling effect T2	135	- 3.279	3.577	0.067	1.035	53.87	53.11	not sig
bonding effect T2	135	- 2.542	4.027	- 0.018	0.954			
						time 1	time 3	
caring effect at T3	116	- 3.124	2.130	- 0.038	0.840	50.70	51.10	not sig
over -controlling effect T3	116	- 2.676	2.991	- 0.067	0.991	53.39	53.69	not sig
bonding effect T3	116	- 2.758	1.756	- 0.052	0.741			

The overall bonding effect is the average of the caring effect size and the over controlling effect size. The caring and over-controlling concepts refer to the two scales on the Parenting Bonding Instrument (PBI), a self report instrument. The raw scores were converted to standard scores based on the standard deviation of a normative sample reported in the literature of 843 Israeli students (Canetti et al, 1997). The norms vary according to whether parent being rated is male or female.

7.4-a Interpretation of Bonding Effect

Of the 135 children in Bayfield who were administered the PBI, the mean standard score was almost exactly equal to the normative sample mean. The effect size for the Bayfield cohort was zero. The standard deviation is very significant– varying from 14 times the effect mean to 50+ times the effect mean. Clearly, this variation means profound variations by sub groups of children.

The variation on both clinical and bonding scores means Bayfield's resident population is quite heterogeneous from a clinical point of view. For this reason, we examined the effect size by distinct groups of children, using the clinical and bonding scales as grouping variables.

8.0 Analysis of Change by Sub-group

The data analysis is Section 7.0, Analysis of Change, was based on a statistic referred to as *the effect size*. This is a very important statistic because it allows for direct comparisons with other studies in the literature especially meta-analyses. The meta-analysis on young offenders (Dowden & Andrews, 2000) shows a consistent, small effect when treating young offenders, even with the "most promising interventions". This study bases comparisons across different interventions with a measure of change that is standardized and directly comparable regardless of the numbering system used in the measurement tools.

The effect sizes for children in Bayfield vary significantly, measurement tool by measurement tool. We know from Section 6.3, Table 3, that the different measurement tools are not redundant; they measure different concepts. The varying effect sizes across different measurement tools mean that children in Bayfield change in a positive direction in some respects and in a negative direction in other respects.

Secondly, the effect sizes, themselves vary from zero effect to small and moderate depending on what domain is being measured. This suggests that the educational and treatment programs at Bayfield are more successfully in some domains than in others.

Thirdly, the standard deviation of the effect size is very large (ranging from 14 times to 53 times the mean for attachment scales). The degree of variance suggests that Bayfield is treating a heterogeneous population, some of whom respond much better to the interventions than others. In view of this conclusion, we split the dataset into sub-groups of children based on the pattern of scores in the major domains (attachment, education, clinical and stress)

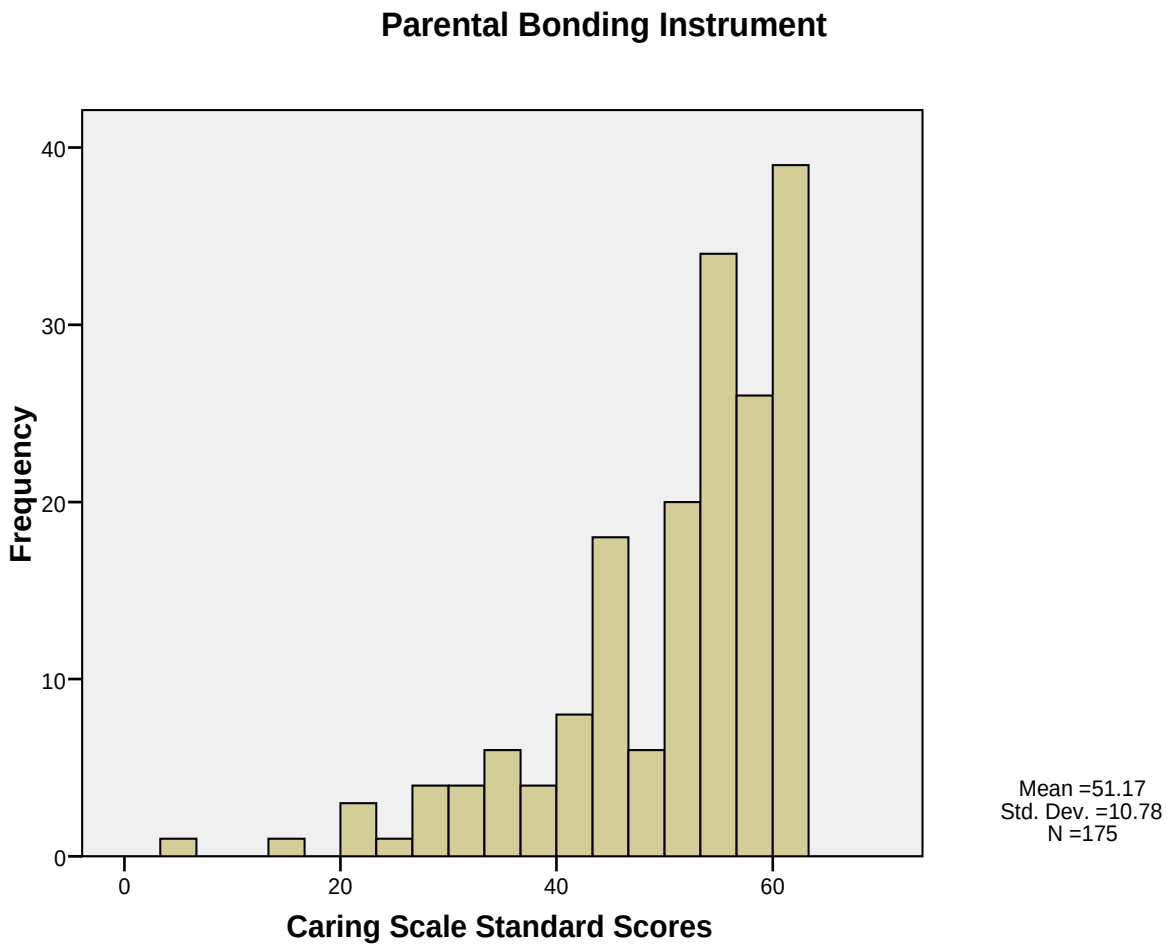
8.1 Attachment

The concept of attachment has a central place in the theoretical model of change which guides the program logic model and the management priorities of Bayfield. The measurement tool for dealing with attachment is the Parental Bonding Instrument which is well established in the literature, the most frequently used and validated tool, for assessing attachment in teenagers and adults.

8.1-a Caring Scale

The histogram (Figure 1) shows the pattern of caring scores at T1.

FIGURE 1: THE CARING STANDARD SCORES AT T1



Seventy-seven percent of the children in Bayfield report that they feel cared about by the *parent figure they are closest to*, resulting in standard scores that are in the range of 45 to 62. About 20% of these children were "faking good"; they reported that they felt cared about at T1 and nine months later reported that they felt the opposite with standard scores substantially below 45. These children appear to have deteriorated, with a strong negative effect score at T2 (-1.438). These two separate forces (a) children whose scores deteriorate at T2 and (b) those whose scores were already in the positive range at T1 and have little room to improve, bring down the overall effect score for the caring scale (see Section 7.4-a). This also explains why the standard deviation of effect scores for the bonding scale is so large.

TABLE 11: CHILDREN WHO FEEL PROFOUNDLY UNLOVED AT T1

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Care standard scores at T1	40	5.465	44.735	34.959	9.092
Care standard scores at T2	28	16.711	62.434	47.134	13.006
Caring effect size at T2	28	- 1.894	4.477	**1.167**	1.453
CGAS effect at T2	32	- 2.712	1.964	- **0.532**	1.043
CGAS effect at T3	24	- 2.805	1.870	- **0.627**	1.167
SA_effect1	23	- 2.277	1.386	- **0.405**	0.870
ReadST at T2	10	- 0.469	1.501	**0.324**	0.634

There are 40 children, or 23% of those tested on the PBI at T1, whose standard scores at T1 fell below 45. The average "standard score at T1" seen in row 1 of Table 11 is 35. By T2, these children had a mean score of 47, within the normal range, a phenomenal improvement. The caring effect size (row 3) for this group of children is very strong, 1.167. These children produced moderate effect sizes in other clinical scales (CGAS and SA-45). Finally, this group improved significantly in their standard reading scores (effect size of .324) indicating that they were catching up to their peers in reading skills. There are two potential threats to the truth in this table: (a) regression from the mean and (b) the number of cases is too small. Regression from the mean refers to a common pattern that very high (or very low) scores tend to converge toward the mean over time. This might explain the strong positive effect on the caring scale when the data is restricted to children who scored very poorly at T1. However, the regression from the mean argument does not explain the moderate-strong positive effects in the CGAS, SA-45 and reading scales, which were all in the Bayfield average range at T1. The strong relationship between improvements in the caring scale and other independent measures[4] supports the view that changes in the caring scale represented a substantive change in the individual. However, the only conclusive counter-argument to the regression from the mean explanation is to use a *no-intervention control group*.

The second threat comes from the small number of cases; there are only 28 cases in this group had a T2 score and only 10 of these cases were tested on the WIAT. In the ideal research environment, children would be selected for admission based on homogeneity of the control variables.

[4] note: their independence was proven in Section 6.3 by factor analysis

8.1-b Over Controlling Scale

The over controlling scale of the PBI measures the degree to which the child feels "fairly treated and reasonably supervised". Approximately one half, 70 of the 135 youth who were tested at T1, scored above 60 on this scale, meaning that they felt they were treated unfairly by the parent they referenced in the test. Young teenagers, in particular, struggle with issues of authority and control. This is a core attachment challenge of school age children and early adolescence (Crittenden, 2000; Waters et al, 1991; Richters and Waters, 1990). If teenagers can change their attitude about authority and what parental behaviour constitutes fair treatment and reasonable control, they can navigate all of their relationships more effectively. This is a central element in the program logic model of Bayfield.

The chart demonstrates that Bayfield is able to help teenagers who feel unfairly treated to change their inner working model of interpersonal relationships. Specifically, the effect size on the over controlling scale is higher than moderate (.678) after only nine months of treatment. This sub group of children also showed moderate effects in their functional adaptation (CGAS at -.340) and in the number of psychiatric symptoms (SA-45 at -.440). This result is similar to the children who felt unloved. The two scales of the PBI are correlated, but distinct domains. About one third of the children who feel unfairly treated also feel unloved (24 out of 70). It appears that each sub-group, unloved and unfairly treated, affect other clinical scales in the same way.

The two scales have a distinct effect on academic achievement. The children who feel *unloved*, not only get better in terms of this feeling, but their reading is significantly improved to a moderate degree.

The children who felt *unfairly treated by their parents*, not only improved on this feeling, but their math and oral learning also improved as reflected in the last three rows of the table below.

TABLE 12: CHILDREN WHO FEEL UNFAIRLY TREATED BY THEIR PARENTS AT T1

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Over controlling at T1	70	55.736	94.958	64.198	9.289
Over controlling at T2	56	39.309	91.585	56.421	11.305
over control effect at T2	56	- 1.994	3.577	**0.678**	1.045
CGAS effect at T2	63	- 2.712	1.964	- **0.340**	1.043
SA_45 effect at T1	43	- 2.277	1.386	- **0.440**	0.856
ReadST effect at T2	18	- 0.985	1.595	0.258	0.771
OralST effect at T2	18	- 0.589	1.607	**0.336**	0.565
MathST effect at T2	18	- 0.327	1.362	**0.318**	0.478
Average Academic effect T2	18	- 0.329	1.521	**0.304**	0.428

It should be noted that the PBI testing was done before the academic testing. The first pre-post period, T2, on the PBI predates the first pre-post period in academics (T2). There are three sine qua non criteria to prove causality:

(a) The causal agent must precede the "effect"

(b) The causal agent and the effect must be correlated (the stronger the better)

(c) The correlation must not spurious: i.e., driven by another variable that is correlated with the other two (the supposed cause and the effect)

The first two criteria for showing a causal relationship between improvements on attachment and subsequent improvements in reading, oral learning and math have been met. Usually, the third criterion is met though an experimental design which controls for all other possible explanatory factors.

There are three problems with the data analysis to date. First of all, it is possible that some other domain measured by Bayfield (psychiatric symptoms, hyperactivity, functional ability or aspects of their background) explains the academic progress as well or better than the attachment scales. Secondly, a correlation between attachment and academics does not measure how much of the variance in academics is "explained" by changes in attachment. Thirdly, a factor outside the scope of the research (genetics, brain injury, instructional method) may explain some portion of the variance in academic change. Answering these questions can be estimated by means of a factor analysis, as well as multi-factor regression analysis.

A factor analysis of the effect sizes, adversity (physical and sexual abuse) and baseline scores was completed. The factor analysis identified the underlying structure of the relationships between the variables. The tables below answer some of the questions listed above:

TABLE 13: FACTOR 3: READING, MATH AND BONDING

Responders in attachment, reading and math scores	**13%**
	Factor loadings
ReadST effect size at T3 (i.e. good outcomes in reading at T3)	0.710
Do not have a history of sexual abuse	0.667
Over controlling effect size at T2	0.658
Very low caring standard score at T1	- 0.649
ReadST effect size at T2 (good outcomes in reading at T2)	0.627
CGAS – effect size at T3 (good outcomes in CGAS at T3)	- 0.615
MathST effect size at T2 (good outcomes in math at T2)	0.553
Caring effect size at T3 (good outcomes in caring scores at T3)	0.531

Factor 3 can be thought of a group of characteristics that are highly correlated with each other but are distinct and uncorrelated with other hidden factors. The analysis also found that 13% of the variance in all of the scores (academics, clinical, stress related and attachment) can be explained by factor 3. This result supports the Bayfield theoretical explanatory model (improvements in attachment leads to improvement in academics), but other factors are also important.

The factor loadings are the same statistic as path co-efficients in a causal model (SPSS version 15.0, manual). The factor analysis was conducted with 50 variables. All but eight were dropped out by SPSS based on covariance matrix.

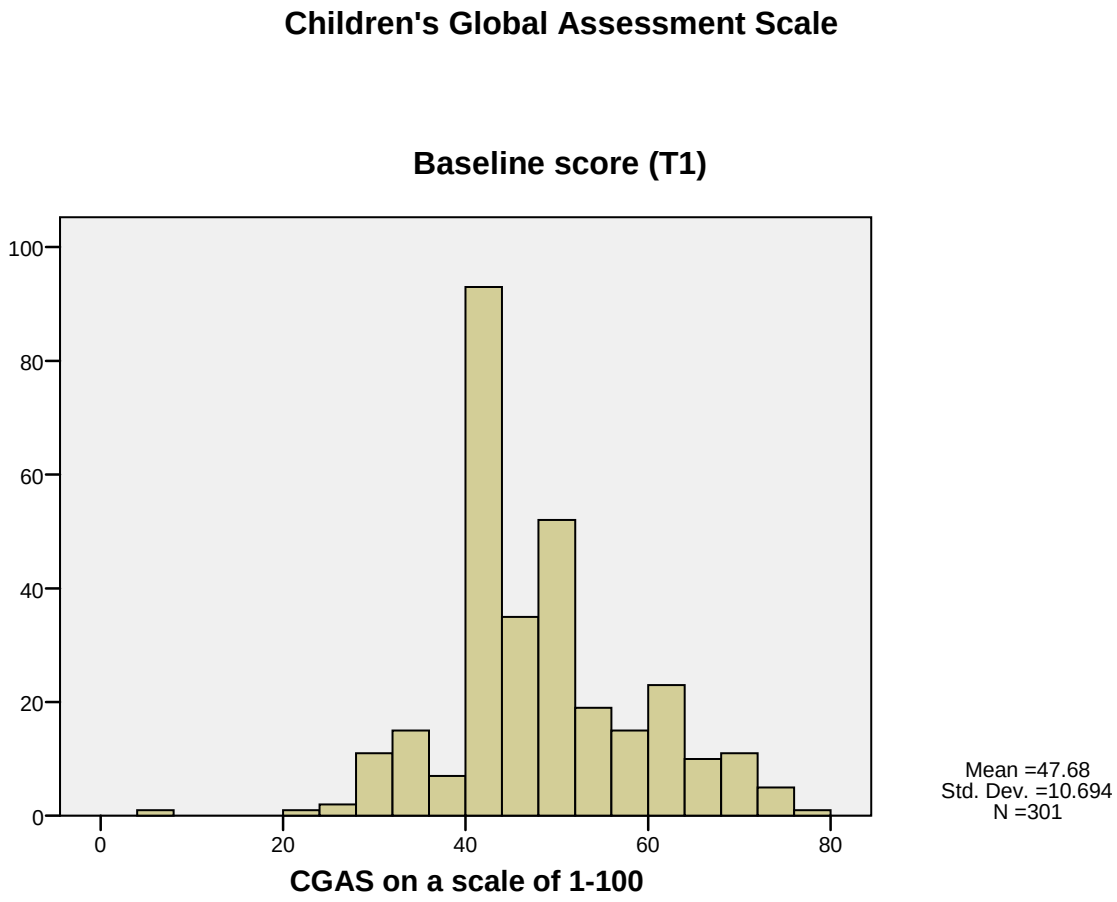
8.2 Ability to Cope with Society

The Children's Global Assessment Scale measures how well the child adapts to society and how well he carries out the major roles of his life at home, school and in the community. The CGAS is axis V of the DSM-IV; it has been found in longitudinal designs to predict future adjustment better than the Axis I diagnosis.

The histogram shows that the vast majority of residents fall between 40 and 60 on the CGAS. The following text describes children in the 40-60 range:

60-51	*Variable functioning with sporadic difficulties or symptoms in several but not all social areas*; disturbances would be apparent to those who encounter the child in a dysfunctional setting or time but not to those who see the child in other settings
50-41	*Moderate degree of interference in functioning in most social areas or severe impairment in functioning in one area*, such as might result from, for example suicidal preoccupations and ruminations, school refusal and other forms of anxiety, obsessive rituals, major conversion symptoms, frequent anxiety attacks, poor or inappropriate social skills, frequent episodes of aggressive or other antisocial behaviour with some preservation of meaningful social relationships

FIGURE 2: BASELINE SCORES OF THE CGAS: ALL CHILDREN IN BAYFIELD



The true scale ends at 100. This means that there is lots of room for the children to improve. By T2, there is a small to moderate effect size (-.376) on this measure. We isolated the children who were most dysfunctional (scores under 40) in order to assess how the most dysfunctional group responded to intervention:

TABLE 14: DYSFUNCTIONAL CHILDREN

						mean standard scores		
Descriptive Statistics	N	Min effect	Max effect	Mean effect	Std. Deviation	time 1	time 2/3	Diff t/sig
CGAS effect at T2	63	- 3.366	0.561	- **0.990**	0.992	35.90	46.49	- 7.9
CGAS effect at T3	54	- 3.834	0.187	- **1.101**	0.919	35.78	47.56	- 8.8
oralST1_2	33	- 1.178	1.981	**0.406**	0.652			
oralST1_3	25	- 1.071	1.874	**0.559**	0.659			

The most dysfunctional children are not more likely than any others to show special problems with attachment; the caring and over controlling scores are very close to 50 for this group. This outcome demonstrates that attachment scores at T1 are independent of the clinical scores. By T2, about nine months after admission, these children have improved significantly on the CGAS reflected in the very strong effect size (-.990). In addition, these children are associated with a moderate effect in oral learning standard scores (.406).

The factor analysis identified the scores on the CGAS as the defining element in Factor 1.

TABLE 15: FACTOR 1, SOCIAL COMPETENT CHILDREN, WITH LOTS OF EARLY ADVERSITY WHO IMPROVE ON THE OVER-CONTROLLING SCALE AND IMPROVE IN ORAL LEARNING

higher baseline scores on CGAS, adversity, good oral learning and positive bonding	21%

	Factor loadings
CGAS at T1: favoring those with higher scores	0.923
Socio-demographic checklist: favoring higher scores	0.877
Associated with a teenage mom (parent of resident)	- 0.866
SA effect at T2: not positive, but not in clinical range at T1 either	0.789
LOA at T3: the amount of adult support needed – positive outcomes at T3	- 0.676
SA effect at T3: not positive see above	0.648
Over-control effect at T3: positive outcomes on the over controlling scale at T3	- 0.648
OralSt at T2: positive outcomes in oral learning	0.644
SA45: associated with scores in the normal range	- 0.626
OS effect size at T3: more distress in current issues at T3	0.592
CGAS effect size at T2: poorer outcomes, but they are starting high at T1	0.585
CGAS effect size at T3: poorer outcomes	0.569

This factor explains 21% of the variance in test scores. Embedded in this factor is the positive correlation between improvements in the over controlling scale and oral learning; this relationship appears frequently in different statistical methods. However, in this analysis, the oral learning preceded the improvements in the adolescent's sense of fairness (over-controlling scale). This means that different groups of children show a different order of change.

The roles of cause and effect can reverse from one group to another. Although, we cannot prove this through this data analysis, the literature suggests that cause and effect is an iterative process in psychotherapy.

8.3 Level of Care (Adult Support in Social Domain)

The following is the LOASocial scale.

SOCIAL FUNCTIONING	SPECIFIC SKILLS OR ACTIONS	SCORE
Initiates	joins an ongoing interaction or starts a new one	
self regulates	Manages own behaviour without instruction from others	
follow rules	Follows rules, guidelines and routines of activities	
provides positive feedback	Provides positive feedback & reinforcement to others	
provides negative feedback	Provides negative feedback or consequences to others	
obtain cues	Obtains and responds to relevant situational cues	
provides information/offers assistance	To others	
requests/accepts assistance	From others	
indicates preference	Makes choices from available alternatives	
cope with negatives	Exhibits alternative strategies to cope with negative events	
terminates	Terminates or withdraws from an interaction &/or activity	
Self Care and home living	Eating	
	Grooming and dressing	
	Washing hands	
	Toileting	
	Taking a bath or shower	
	Cleaning up after an activity	
	Identifying physical needs, such as elimination or hunger	
General community functioning	Going from place to place in the community	
	Crossing street safely	
	Attending Community School	
	Using Community Recreation Facilities	
	Total support score for social functioning skills = sum of individual scores	

The numbers range from 0 to 4 based on the following criteria:

LEVEL OF CARE	SPECIFICATION	SCORE
independent	no special needs compared to children of their age	0
daily oversight	Child requires daily supervision	1
verbal prompt	Child requires verbal prompting	2
physical prompt	Child requires hand over hand guidance	3
total care	Staff must do everything	4

There is a maximum of 88 points and a minimum of 0. Only 1% of children in Bayfield have a score of 0 on admission. About one quarter (23%) score below 14 points; we used a score equal or above 14 to identify "needy" children.

There is no difference between the baseline scores in attachment, education and clinical scales for all measurements except the LOASocial. Therefore, we should not expect to see any difference in effect sizes unless the LOASocial is interacting with outcomes. There are differences in effect sizes throughout as illustrated below:

TABLE 16: CHILDREN NEEDING MORE ADULT SUPPORT IN SOCIAL INTERACTION

needy kids					
	N	Minimum	Maximum	Mean	mean for all children
LOAsocial	169	15.000	59.000	26.822	22.695
LOA effect at T2	146	- 3.360	2.387	- 0.187	0.006
LOA effect at T3	104	- 3.625	2.741	- 0.240	- 0.020
LOA effect at T4	72	- 3.890	2.564	- 0.270	- 0.084
ReadSt baseline score	88	40.000	121.000	76.511	77.158
ReadST effect at T2	88	- 1.829	1.876	0.200	0.147
ReadST effect at T3	55	- 1.970	1.970	0.294	0.241
MathSt1	90	40.000	130.000	70.522	71.405
MathST effect at T2	89	- 0.872	1.471	0.129	0.077
MathST effect at T3	56	- 1.526	2.507	0.074	0.055
OralSt1	90	40.000	136.000	83.056	82.819
OralST effect at T2	88	- 1.178	2.731	0.283	0.247
OralST effect at T3	56	- 1.071	1.874	0.440	0.448
Mean academic effect T2	89	- 0.757	1.521	0.207	0.161
Mean academic effect T3	56	- 0.604	1.675	0.269	0.246
CGAS	169	6.000	74.000	45.361	47.681
CGAS effect at T2	151	- 3.366	2.712	- 0.300	- 0.376
CGAS effect at T3	112	- 3.647	3.927	- 0.518	- 0.490
SA45	76	36.000	74.000	59.145	58.677

needy kids					
	N	Minimum	Maximum	Mean	mean for all children
SA effect at T2	52	- 2.277	1.782	- 0.324	- 0.285
SA effect at T3	40	- 1.782	2.079	- 0.438	- 0.340
SA effect at T4	27	- 2.872	2.079	- 0.708	- 0.559

OS	162	-	14.000	3.494	3.788
OS effect at T2	133	- 3.006	5.410	0.253	0.147
OS effect at T3	97	- 2.705	3.306	0.118	0.184
OS effect at T4	64	- 2.104	4.208	0.338	0.261

8.3-a Interpretation of Outcomes for Needy Children

The cut-score for identifying "needy" children was 14. More than 75% of the children are at or above this cut score. At this level, the LOASocial is showing a small treatment effect, compared to no treatment effect, when 100% of children are considered.

Reading, math and oral learning all show stronger effects, even though the baseline scores are virtually the same. Clearly "needing" more adult support is correlated with better outcomes in school. This finding provides indirect support for the hypothesis:

> *Children, who need more support to manage social interaction, get more support and as a consequence receive more supportive interactions with staff; supportive interactions help children to feel cared about by people in their immediate environment. Children who feel cared about to a greater degree learn more.*

The effect sizes related to psychiatric symptoms (SA-45) are much better and continue through four separate waves of testing. The CGAS effect score is greater at T3 compared to the "all children group". Finally, the Objective Stressors Checklist, which measures the number of issues/concerns that children currently experience, is showing better effects with this group than with the total population – especially at T3.

8.3-b Factor Analysis: The Hidden Factor in the LOA

The second factor derived from the factor analysis explains 16% of the variation. The three factors discussed so far explain 40% of the variation in effect sizes. The hidden factors can be thought as discrete groups of children, since the factors are derived from data elements that are lined up across the columns related to each unique child.

In Section 6.3 the domains of academics, clinical, attachment and stress are discrete factors. The baseline scores line up directly under each of these headings. It is

very significant that these domains cross over and group together when the change effect is considered. This means the relationships between variables emerge over time, exactly what you would see in a causal model.

TABLE 17: FACTOR 2: NEEDY CHILDREN DO BETTER

Children needing a lot of adult support in social interaction, who improve on this variable, also improve in reading, math and oral learning. These children have higher scores on over controlling scale and family members in jail	% of var exp = 16% Varimax rotation Factor loadings
LOA effect size at T2: needing less adult support by T2	- 0.852
LOA effect size at T3: needing even less adult support by T3	- 0.823
LOAsocial: needing a lot of adult support at T1	0.762
ReadSt effect size at T2: better scores	0.741
OralSt effect size at T2: better scores	0.736
Associated with "family member has a history in jail"	- 0.723
MathSt effect size at T2: better scores	0.620
Over-controlling scores at T1: associated with higher scores	0.598

Academic improvement is directly related to good outcomes in the level of adult support needed by children. The factor loading for academics are higher under this factor than under any other, indicating that they correlate most strongly with this hidden factor. Children with higher over controlling scores at T1 are also tied into this group.

8.4 Hyperactivity

The Effect scores related to hyperactivity has emerged as a consistent problem. It appears that a substantial number of children are deteriorating relative to this measure as they get better on the others. This could be an iatrogenic effect: an unintended outcome of the treatment itself. The GCI measures behaviour such as hyperactivity, inattentiveness and emotional liability; it is very sensitive to short term distress and is a good candidate for showing a iatrogenic effect if in fact there is one.

The table below shows the descriptive statistics for the sub group of children who are clinically hyperactive (whose baseline T scores at T1 were above 70 on the Conner Global Index), with the means of all children in the right hand column.

TABLE 18: CLINICALLY HYPERACTIVE CHILDREN

clinically hyperactive children	N	Minimum	Maximum	Mean	Std. Deviation	mean of all children	difference in means
CGI baseline scores at T1	122	70.00	90.00	**79.78**	6.60	65.60	14.18
CGI effect size at T2	105	- 2.480	1.459	**- 0.273**	0.933	0.329	- 0.602
CGI effect size at T3	78	- 2.480	1.459	**- 0.351**	1.031	0.362	- 0.713

clinically hyperactive children	N	Minimum	Maximum	Mean	Std. Deviation	mean of all children	difference in means
CGAS	120	6.00	69.00	42.89	9.23	47.68	- 4.79
CGAS effect size at T2	104	- 3.366	1.870	**- 0.533**	1.065	- 0.376	- 0.157
CGAS effect size at T3	77	- 3.834	2.618	**- 0.605**	1.244	- 0.490	- 0.115
SA45	66	36.00	93.00	61.41	11.61	58.68	2.73
SA effect at T2	43	- 1.881	1.584	- 0.233	0.801	- 0.285	0.052
LOAsocial	77	-	59.00	28.14	11.60	22.70	5.45
LOA effect size at T2	67	- 3.360	1.768	- 0.313	0.948	0.006	- 0.318
caring baseline at T1	73	16.62	62.43	51.84	11.28	51.17	0.67
care effect size at T2	54	- 3.788	3.279	- 0.135	1.445	- 0.092	- 0.043
Over-controlling baseline at T1	73	32.56	94.96	55.91	11.85	53.63	2.28
over-controlling effect size at T2	55	- 1.937	2.617	**0.255**	0.972	0.067	0.188
SD	122	1.0	9.0	4.2	1.8	4.1	0.02
Objective Stress baseline	113	-	12.0	4.1	3.3	3.8	0.34
OS effect size at T2	86	- 2.104	2.705	0.049	0.984	0.147	- 0.098
ReadST effect size at T2	47	- 1.407	1.876	0.172	0.679	0.147	0.024
OralST effect size at T3	48	- 1.178	2.731	0.276	0.650	0.247	0.028
MathST effect size at T3	48	- 0.708	1.471	0.087	0.485	0.077	0.010
ReadST effect size at T3	34	- 1.970	1.970	0.170	0.863	0.241	- 0.071
OralST effect size at T3	33	- 1.071	1.446	0.368	0.582	0.448	- 0.080
MathST effect size at T3	33	- 1.526	2.507	0.035	0.746	0.055	- 0.021

Children whose baseline scores on the CGI are in the clinically hyperactive range display a small effect size. This is much better than the moderate negative effect for the total population. This demonstrates that the negative treatment effects measured by the Global Assessment Scale relates to children who are not hyperactive at T1.

The hyperactive children (at baseline) also show a moderate treatment effect on the CGAS that is quite a bit greater than the treatment effects for the total population. In a complementary finding, these children have a small treatment effect on the Level of Care scale as well compared to no effect for the total population.

There is almost no difference between the hyperactive and non-hyperactive children on the attachment and academic scales.

TABLE 19: FACTOR 4: GOOD OUTCOMES IN ORAL LEARNING AND NON-HYPERACTIVE CHILDREN

Oral learning associated with hyperactivity in a paradoxical fashion	11%

A poor outcome in terms of the CGI effect at T2 : more hyperactivity by T2	0.949
CGI at T1: the baseline scores of hyperactivity are lower	- 0.692
Over-controlling baseline scores at T1 are higher	0.668
Oral ST effect size at T2	0.621
Associated with a history of physical abuse	- 0.510

On factor analysis, the children whose baseline scores on hyperactivity are lower emerged as the distinct group. Noted on Table 18, this group has a small negative treatment effect; they get more hyperactive over time. Nevertheless, they show a positive treatment effect on oral learning.

The other factors which emerged all represent a single domain, which shows some of the variation in scores, especially on academics. This may be due to other variables not considered in this study, e.g., instructional method, genetics, intelligence, the personality of the teacher, the culture in the classroom, may be contributing to the academic outcome.

9.0 Conclusion:

9.1 Findings

9.1-a Baseline Scores

(1) Although the standardized educational testing has been used since the spring of 2005, the results can be generalized to several years of service to the residents of Bayfield. This finding is based on the observation that there are *no materially significant differences (in clinical status, stressors or attachment patterns)* between the group of children treated in the period 2005-2007 and the group treated in the period 1999-2005.

(2) A factor analysis of all test scores at T1 (the first wave of testing) found that there are four latent variables which are distinct and uncorrelated with others. These latent variables are labelled: academic indicators, clinical indicators, measures of stress and measures of attachment.

(3) An analysis of the relationships at T1 produced the following unexpected findings:

a. children who are physically or sexually abused have many more daily hassles and "objective stressors" which they identify in the "here and now" of the program. children who have a family members that was in jail or was raped also report more current distress (Objective stressors)

b. three factors in the background (family member in jail, family member raped and child has a history of abusing drugs) are associated with higher academic functioning that is close to the average functioning of their peers in society.

(4) Children admitted to Bayfield at 13 years of age are more than 2.2 grade levels behind their peers in all academic areas. The explanatory factors include: learning disability, Aboriginal identity, severe environmental stressors including years of poverty, child abuse, and parental substance abuse.

9.1-b Changes over Time

(5) At T2, there is a small effect (.247) in changes on the oral standard score. This means that the children are learning at a faster pace than their peers in the normative sample and they are catching up to the norms in society.

(6) Conversely, there is no effect for changes in the reading and math standard scores at T2. This means that the children are learning new material in reading and math but they are just keeping pace with their peers in society. However, just *keeping pace*, in effect means they have reversed a multi-year negative trend.

(7) By T3, there is a small reading effect (0.241) and a medium oral learning effect (.448). This means that the children in Bayfield are closing the gap with their peers in reading at the rate of 24% every 18 months. They are closing the gap in oral learning ability by 45% in the same period.

(8) Children become more hyperactive and inattentive over time as measured by the Conners' Global Index; the mean effect size is "positive" meaning an increase in symptoms.

(9) Children become more functional; they fit into society better as measured by the Children's Global Assessment Scale and the effect size is in the moderate range. This is very good outcome, clinically, since functional adaptation is the best predictor of future success (Rutter & Sandberg, 1985).

(10) The level of assistance scale is not changing at T2 or T3; this scale measures how much adult support is required to assist the child in the social dimensions daily living. The standard deviation of the LOASocial scale is considerable indicating some groups of children are improving and other groups are not. This is a logical outcome; if the adult assistance required is due to brain injury, e.g., Fetal Alcohol Spectrum Disorder, not much change is expected in this scale.

(11) The SA-45 shows a small improvement at both T2 and T3 that strengthens over time. This improvement is statistically significant as well. However, the test for the 2-wave cohort was administered to 119 teenagers across the larger population (almost every resident in Bayfield over 13 years of age); the baseline score is within the average range of the normative population (t-score of 58.86) well below the clinical range of 70+; the baseline score is close to one standard deviation above the normative mean. This means that the teenagers were reporting a fair number of symptoms (anger, depression, self doubts, and hostility) but these symptoms were not as numerous as one would see in a population of teens in a psychiatric hospital. Nevertheless, these teenagers did make substantial improvement, but the effect size is still small because there is *not much room for improvement* in a population of teenagers who are already within the average range relative to a normative sample.

(12) The FAB-C shows a small improvement at both T2 and T3; almost every child at Bayfield who was too young for the SA-45 was administered the FAB-C as an alternative test; no resident was administered both the FAB-C and the SA-45. The letters in the FAB-C stand for *feelings*, *attitudes* and *behaviour* of children. The FAB-C is a self-report instrument. The FAB-C effect size shows an even greater standard deviation than the SA 45. This means that some children are improving much more than others groups. The baseline score for the entire group of children being administered the FAB-C (n = 82) is 58.27 within one standard deviation of the mean score for the normative sample. There is not much room for improvement when the respondent is this close to the normative mean; the small effect size is to be expected and reflects a positive change in feelings and attitudes overall, and is a good clinical indicator.

(13) Seventy-seven percent (77%) of the children in Bayfield report that they feel cared about by the *parent figure they are closest to*, resulting in standard scores that are in the range of 45 to 62. About 20% of these children were "faking good"; they reported that they felt cared about at T1 and nine months later reported that they felt the opposite with standard scores substantially below 45. These children appear to be less attached to the parent figure they reference in the test, with a strong negative effect score at T2 (-1.438).

(14) There are 40 children (or 23% of those tested on the PBI at T1) whose standard scores at T1 fell below 45, meaning that they felt profoundly unloved by the parent figure they closest to. By T2, these children had a mean score of 47, within the normal range, a phenomenal improvement. The caring effect size

for this group of children is very strong, 1.167. These children also produced moderate effect sizes in other clinical scales (CGAS and SA 45). Finally, this group improved significantly in their standard reading scores (effect size of .324) indicating that they were catching up to their peers in reading skills at the rate of 32% every nine months.

(15) The most dysfunctional children are not more likely than any others to show unique problems with attachment; the caring and over controlling scores are very close to 50 for this group. This demonstrates that attachment scores at T1 are independent of the clinical scores. By T2, about nine months after admission, these children have improved significantly on the CGAS reflected in the very strong effect size (-.990). In addition, these children are associated with a moderate effect in oral learning standard scores (.406).

9.1-c Factor Analysis of Effect Size

In order to identify patterns in the data related to different groups of children and differential response to treatment, a factor analysis of all effect sizes was done.

(16) FACTOR 1: explains 21% of the variance in outcomes, are a subgroup of children which could be described as *social competent children, with lots of early adversity* who improve on the over-controlling scale and improve in oral learning.

These high functioning children also have healthier scores in terms of clinical status at T1; over time they come to appreciate that their parents were not as unfair as they originally felt. These children are not associated with big gains clinically because their lower levels of pathology do not give them much room to improve. These children are the opposite of Factor 2.

(17) FACTOR 2: explains 16% of variance in outcomes are *children needing a lot of adult support in social interaction* who improve on this variable, also improve in reading, math and oral learning. These children have higher scores on over controlling scale and have family members in jail.

Factor 2 was further supported by identifying a sub-group of residents whose score on the LOASocial was above 14 indicating needy children requiring lots of adult mediation to get through the day. These children constitute 77% of the population in Bayfield. The results from examining this subgroup and the associated factor loadings support the following inference:

Children, who need more support to manage social interaction, get more support and as a consequence receive more supportive interactions with staff; supportive interactions help children to feel cared about by people in their immediate environment. Children who feel cared about to a greater degree learn more.

(18) FACTOR 3: explains 13% of the variance in outcomes can be described as *children who are responders in attachment, reading and math scores*.

This result supports the Bayfield theoretical treatment model; *improvements in attachment leads to improvement in academics*, but other factors are also important. It appears children who feel that they are treated very unfairly by their parents and also feel unloved in life, change their feeling of fairness first. This sequence provides an opportunity for change elsewhere. A year later these children also feel that they are loved to a greater degree than they felt on admission. This group are extreme cases of children with profound attachment issues.

This subgroup of children show very strong effect size indicating improvements in the attachment status; this precedes other improvements clinically and academically. To complicate matters, the children continue to improve on attachment measures at T3 and T4, illustrating that real therapeutic change is an iterative and re-enforcing process, not a linear process.

Other patterns, identified in factors 2 and 4, suggest that the same therapeutic process is at work in positive ways with other groups of children, but the Bayfield explanatory model is not as clearly evident, but is supported by inference and reasonable theory.

(19) FACTOR 4: explains 11% of the variance in outcomes, could be described as *Oral learning associated with hyperactivity in a paradoxical fashion.*

This finding suggests that hyperactivity and inattentiveness has a negative effect on oral learning. When all children in the clinical range on the CGI at T1 are grouped together, a different pattern emerges. These children do improve to a small degree and their scores on the over controlling scale of the PBI also improve, as does oral learning. This finding suggests that improving the hyperactivity measure has a positive effect. Unfortunately, many children do not improve on this measure and some get worse, which leads to the pattern identified in Factor 4.

The effect scores related to hyperactivity has emerged as a consistent problem. It appears that a substantial number of children are deteriorating relative to this measure as they get better on the others. This could be an iatrogenic effect: an unintended negative outcome of the treatment itself. The GCI measures behaviour such as hyperactivity, inattentiveness and emotional liability; it is very sensitive to short term distress and is a good candidate for showing an iatrogenic effect, if indeed there is one.

9.2 Confirmatory Analysis

In Section 9.2, we presented the results of multi-factor analysis procedures that appear to confirm the explanation of change that is central to Bayfield's logic model. This confirmatory analysis, however, cannot rule out other possible theories causing the improvement. The competing theories can be ruled out only with new testing or experimental research designs.

We ran a series of multi-factor regression analysis, using the effect sizes for reading, math and oral learning at T2 as the dependent variable and the effect sizes for caring, over controlling, learning disability, CGI, CGAS and LOASocial as the independent variables. The purpose was to identify significant interactions between the variables, which support the Bayfield treatment model and the findings in this paper to this point. This phase of statistical analysis could be described as confirmatory analysis.

9.3 Discussion

A draft of the paper was circulated to the Bayfield Clinical Team and Leadership Team followed two weeks later by a focus group to discuss the findings. A dialogue concerning the four factors was central to the discussion:

It is suggested that separation assists the memory of unfairness (over-controlling) to fade as observed in the sessions held with the family. The inclusion of the family in the treatment helps the child to see his family from a safe place and subsequently interprets their behaviour differently. Also observed is that the families begin to behave differently toward their child in a therapeutic setting while realizing the changes in their son and themselves. It is important to the child that his parents/family continue to be safe and accessible, and for some children, the possibility of reintegrating into the family helps the child to create a voice of optimism.

As mentioned the average academic lag is 2.2 years behind their public school peers when admitted to Bayfield. The child's history of academic failure appears to this pattern, with lower expectations and an emotional overlay preventing him from experiencing success in the school environment. Strategically the curriculum is geared for the child to succeed and subsequently creates a tipping point as his new sense of identity enables him to experience and predict success. The child develops a new perception of self embedded in feelings of inclusion. This need of belonging and acceptance in a caring environment outweighs his need to be seen by expressing his anger and aggression.

Residentially, staff are pervasively trained to help the child to self reflect through consolidating the combined effort of implementing one theoretical framework or treatment approach. The teams working with each child set the tone for attachment and this research supports that the strategy is resulting in good outcomes.

References

Canetti, L., Bacher, E. Galili-Weisstub, E., Kaplan De-Nour, A. & Shalev, A. (1997), "Parental Bonding and Mental Health in Adolescence", *Adolescence,* 32 (126), Summer, 381-394.

Cohen, J. (1969), *Statistical Power Analysis for the Behavioural Sciences*, N.Y.: Academic Press.

Crittenden, P. (2000), "A Dynamic-Maturational Approach to Continuity and Change in Pattern of Attachment", in *The Organization of Attachment Relationships*, Patricia Crittenden & Angelika Claussen (eds), Cambridge University Press: Cambridge, UK, pages 343-357.

Factor, D & Fulton, R. (1999), Social-demographic Checklist.

Lipsey, M. & Wilson, D. (1993), The Efficacy of Psychological, Educational and Behavioural Treatment, confirmation from Meta-Analysis", *American Psychologist*, 48(12), 1181-1209.

Lipsey, M. (1992), "Juvenile Delinquency Treatment: a meta-analytic inquiry into the Variability of Effects", in T.D. Cook, H. Cooper, D.S. Cordray, H. Hatmann, L.V. Hedges, R.G> Light, T.A. Lewis & F. Mosteller (eds), *Meta-analysis for explanation*, (pp 82-127), N.Y.: Russell Sage Foundation.

Dowden, C. & Andrews, D.A. (2000), "Effective correctional treatment and violent re-offending: a meta-analysis.(Canada)", *Canadian Journal of Criminology*, 42(4), 449.

Powers, K., Potthoff, S. J., Bearinger, L. H. & Resnick, M.D. (2003), TITLE: Does Cultural Programming Improve Educational Outcomes for American Indian youth?

Richters, J. and Waters, E. (1990), "Attachment and socialization. The positive side of social influence", In Feinman, S. and Lewis, M. (Eds), *Social influences and behaviour*, NewYork: Plenum, 185-213.

Rutter, Michael and Seija Sandberg (1985), "Epidemiology of Child psychiatric Disorder: methodological Issues and Some Substantive Findings", *Child Psychiatry and Human Development*, 15(4), 209-233.

Sanders, L. (2003) "Attachment of Adolescent Males in a Residential Treatment Setting," UMI Publication, Ann Arbor, Michigan.

Sanders,L. & Jamieson,J. (2004), " Predictors of Academic Achievement for at Risk Adolescent Males in a Residential Treatment Centre." Fielding Institute, Santa Barbara California.

Sanders,L. & Fulton, R. (2006), "Educational Achievement and Attachment, Bayfield School Outcome Study", Fielding Institute Post Doctoral Seminars, Monterey, CA.

Stuart, C. & Sanders L., (2007), Understanding the Relationship between Client Outcomes, Childhood Resiliency and Quality of Care in Residential Care and Treatment," Ryerson University (2007-2008).

SOURCE: Journal of American Indian Education 42 no2 17-49 2003.

Waters, E., Kondo-lkemura, K., Posada, G. and Richters, J.E. (1991), "Learning love: mechanisms and milestones", In Gunnar, M.. and Sroufe, L.A. (Ed Minnesota Symposium on Child Psychology.Vol. 23. Hilisdale, NJ: Eribaum, 217-55.