ANU/ACT Education Directorate Research Partnership 2019 – 2021

TASK 4 –POSITIVE BEHAVIOUR FOR LEARNING: AN INTERVENTION TO BUILD SCHOOLS CLIMATE AND SCHOOL IDENTIFICATION

AUTHORS: Diana Cárdenas, Eunro Lee & Kate Reynolds

ACT Government
Education and Training

Australian National University

POSITIVE BEHAVIOUR FOR LEARNING: AN INTERVENTION TO BUILD STRONGER SCHOOLS

Schools around the globe are seeing a decline in positive outcomes such as school belonging and school engagement along with increases in mental illness and loneliness (Allen, 2019). This is concerning, given that schools play a key role in forming and shaping students' lives as they transition into adulthood. Fortunately, schools are also well-placed to create and implement interventions to improve students' lives. Recently, many schools in the ACT have adopted a school intervention that aims at improving student outcomes: Positive Behaviour for Learning or PBL. Currently, 36 Primary Schools, seven P-10 schools, nine High Schools, are participating in PBL-based programs. Despite it being increasingly applied, there has been no systematic investigation on the effect of PBL on student outcomes ranging from well-being and mental health, to engagement, to school climate and school identification.

Based on the ACT Education Directorates 2019-2021 strategic plan (https://www.education.act.gov.au/__data/assets/pdf_file/0010/1482463/Strategic-Plan-Our-Priorities-for-2020.pdf), there is a focus on supporting students' learning and well-being. Such efforts need to be informed by data in decision-making so they are evidence-based. Data is essential in monitoring progress and informing policy, service delivery and targeting of resources. In line with this, the School Climate and Satisfaction Survey (which informs the ACT Education Directorate and Australian National University ANU project) is a valid and reliable school-based assessment tool. The ongoing research collaboration and data collection is designed to enable evidence to be embedded in school improvement and to evaluate the impact of school initiatives to strengthen learning and well-being.

In line with this evidence-based emphasis, the current report examines the effectiveness of PBL in promoting students' well-being (indexed by **depression, anxiety and positive emotions**), supporting their engagement in school (assessed as **behavioural engagement, emotional engagement,** as well **support and safety**), and more fundamentally promoting a positive **school climate** and strong **school identification**. To investigate these relationships, we compare six ACT schools who adopted PBL early (in 2015 or 2016) and therefore would produce discernible improvements due to PBL to six schools who have adopted PBL late (after 2017) or have not engaged in this intervention as yet. It also examines whether implementing Positive Behavioural Learning adequately (i.e., high **fidelity**) contributes to these student outcomes. The results of these analyses support PBLs effectiveness, particularly when there is high fidelity. Lastly, we conclude with outlining practical implications to increase PBL's ability to promote positive school outcomes.

School-Wide Positive Behavioral Intervention and Support

Schools are uniquely placed to support young people as they transition into adulthood. The school is recognised as an environment that can impact significantly on students' academic achievement and well-being. In a context where these outcomes are in decline in Australia and elsewhere there is pressure and urgency to identify new ways in which schools can contribute to building positive futures for young people.

Many interventions have been created, designed and employed to this end but often they suffer from being piecemeal, small scale and not assessed using high-quality evidence. Against this backdrop,

there is one framework that has received wide empirical support is School-Wide Positive Behavioral Interventions and Supports (SWPBIS). SWPBIS is a framework that uses empirical data from each specific school and integrates its unique aspects (i.e., its own data, systems and practices) to create "the kinds of schools where all students are successful" (PBIS, 2019). Its main components involve creating clear behavioural expectations for students, teaching these behaviours, reinforcing students that manifest these behaviours, and adjusting interventions based on data from that school (Horner et al., 2009). While every school may have its own unique set of behavioural expectations, these are put forward and promoted in a systematic way via the SWPBIS framework. As a school-wide framework, it is applied consistently to the entire school, across all contexts (courses and free-periods) and students.

SWPBIS has three tiers of interventions, with the first tier being a general (or universal) prevention program employed with all students in school. In this first tier, schools and PBL-leaders present and teach the main rules of the school (which are positively framed; e.g., "Be respectful"; "Be ready to learn") to all students and staff. This allows all students to know the behavioural expectations and the positive and negative consequences of following these expectations. The second and third tiers are more specialized, with interventions aimed at students who are not responsive to the first tier intervention (Horner et al., 2009).

Research of a very high standard consistently shows SWPBIS to have positive outcomes although there are some areas where the quality of the data to assess such programs can be strengthened (see below). One example of research evidence is a randomized wait-list controlled trial of SWPBIS conducted in elementary schools over a 3-year period showed that staff who were in SWPBIS schools perceived greater school support and safety (Horner et al., 2009). In a similar vein, staff from schools who had SWPBIS reported greater organization health and staff relations (Bradshaw et al., 2008). Teachers in schools that had employed PBI also reported less behaviour problems, concentration problems, as well as greater social-emotional functioning and prosocial behaviour in their students (Bradshaw et al. 2012; see also Waasdorp et al., 2012). Schools who had implemented SWPBIS for over 5 years were also less likely to have students suspended and office discipline referrals (Bradshaw et al., 2009). There is also evidence that SWPBIS is most effective when it is well implemented, i.e., when the practical implementation follows closely or is faithful to the original instructions set forward by developers. High **fidelity** of SWPBIS implementation has been associated with decreases in office referrals, in-school suspensions and out of school suspensions (Childs et al, 2010) as well as collegial leadership, organization health academic emphasis as reported by staff (Bradshaw et al., 2008; for synthesis articles, see Estrapala et al., 2020; Noltemeyer et al., 2019).

An issue with the evidence to date is that it is largely staff reported (although some administrative indicators are also used) and collected within the SWPBIS program. There are many reasons why staff at schools that have allocated time and energy to the implementation of SWPBIS may report positive impacts (compared to control schools which have continued business-as-usual). Furthermore, existing work ignores the perspective of students and does not necessarily focus on the critical dimensions of the school environment which are likely to contribute to sustainable improvements in staff and student outcomes. In line with emerging work in education and psychology, where ACT Education Directorate is at the forefront, school climate and school identification are factors that sit at the centre of the impact of schools on staff and students outcomes.

Positive Behaviour Support for Learning in the ACT

SWPBIS has been most widely applied and researched in the USA with very little research outside this context. Nevertheless, SWPBIS is being increasingly adapted and adopted world-wide, including in the ACT where it is named Positive Behaviour for Learning (or PBL). PBL follows the same principles of SWPBIS (ACT Government: Education). It begins with the creation of clear rules by the school community (staff, students, and parents). These rules are positively framed (e.g., "raise your hand" instead of "do not interrupt your teacher"). Following this, behavioural systems are created to both reinforce (and thus make more reoccurring) student behaviours that follow the rules and dissuade behaviours that are not in line with the rules. In addition, data were collected from schools to examine the fidelity of PBL application and pinpoint places in the framework that need reinforcement.

The ACT Education Directorate 2018-2021 strategic plan is to create schools in which students feel safe, supported and empowered to be active in their own education *by supporting students learning and well-being* (emphasis added; ACT Government, Education, 2020). Despite being theoretically aligned, given the 5 year data is only now becoming available there is to date no systematic and robust evaluation to investigate if the PBL framework is successfully supporting this strategic plan. Hence, it would be timely and informative to examine whether the PBL framework improves students' well-being (depression, anxiety and positive emotions), supports their engagement in school (behavioural engagement, emotional engagement, as well support and safety), and importantly, promotes a positive school climate and strong school identification in ACT schools.

Current project

The goal of this research report is to examine whether students in schools that have adopted PBL's Tier 1 Support will have the following:

1. better mental health outcomes (depression, anxiety, positive emotions);

2. greater learning engagement;

3. greater school climate and school identification.

This last question is of particular importance because of the ubiquitous nature of their impact on school outcomes. A research partnership between the ACT Education Directorate and the Research School of Psychology at the Australian National University has consistently demonstrated that a positive school climate and a strong school identification result in a range of positive student outcomes, from lower bullying and victimization (Turner et al., 2018), to higher NAPLAN scores (Maxwell et al., 2017) to better well-being (Bizumic et al., 2009). Given the central role of school climate and school identification in predicting student outcomes, there is a need to examine and identify school interventions that can effectively strengthen these variables, and in turn, improve student outcomes.

The positive effect of PBL is examined in two ways. First, we compare students in schools that have applied Tier 1 support of PBL for several years (now called PBL schools) to students in a matched group of schools that applied PBL late/did not apply it at all (from now on named control schools). Second, we examine whether the fidelity in adopting the critical features of PBL (i.e., high fidelity) is also positively associated with better student outcomes. Importantly, student outcomes are measured as part of an annual school climate and school identification assessment that takes place in

complete independence of the PBL team. Therefore, it offers a rigorously independent test, namely students' unbiased perspective of being (or not) in a PBL school.

Key in providing the highest level of evidence, this project matched schools based on school-size and **ICSEA** so that they were as similar to each other as possible. In addition, it examines the effects of PBL on measures taken from the same individual student over a five-year period (a **longitudinal design).** Thus, this project utilises higher quality data and can better inform the education sector.

Methodology

For details on the composition of the sample and measures employed, see Appendix 1.

The project involved 12 public schools (P-10 and high schools) of the ACT (for a detailed list of schools, see Table 1 in the Appendix). Six of these schools had been applying PBL since 2015 (except for Calwell, which started in 2014). These six schools were matched to six other P-10 and high schools based on school size and ICSEA (see Table 1). We analysed student survey responses from five years ($N_{all\ years}$ = 9952): Year 0 ($n_{year\ 0}$ = 2380) is the baseline year, before PBL schools began applying this intervention (2015 for all schools and 2014 for Calwell and Belconnen high schools); Year 1 ($n_{year\ 1}$ = 2396) is the year in which the intervention began for PBL schools (2016 for most schools); lastly, Years 2 ($n_{year\ 2}$ = 2574; 2017 for most schools), 3 ($n_{year\ 3}$ = 3497; 2018 for most schools) and 4 ($n_{year\ 4}$= 4060; 2019 for most schools) provide important evidence of long term effects of PBL. Students present during scheduled class times (during the third term of each school year) answered the questionnaire online after providing consent.

The students' questionnaire included measures of positive affect, depression, anxiety, behavioural engagement, emotional engagement, safety and support, school climate and school identification as well as their sex and their grade. Student numbers were used to link the longitudinal data from each annual survey round. Data from the education district record (parental education, ICSEA in the first year, staff retention, student enrolment) were included, as was the fidelity with which schools had adopted PBL (the extent to which school personnel are applying the core features of PBL, such as defining the expectations and teaching these expectations).

Key Findings

For details on the statistical analysis and results, see Appendix 2.

The following analyses examined whether students' mental health, learning engagement as well as school climate and school identification progressed differently in time (i.e., across years) depending on whether students went to PBL schools versus control schools. These differences are examined while controlling for important school variables (school ICSEA, staff retention rate and student enrolment) and student-level variables (parental education, grade at year 0 and sex). Overall, results of mixed ANOVAS suggest that students in PBL had overall better student school outcomes than students in control groups (including lower depression, greater positive affect, greater emotional engagement, greater behavioural engagement, greater support and safety, and greater school identification).

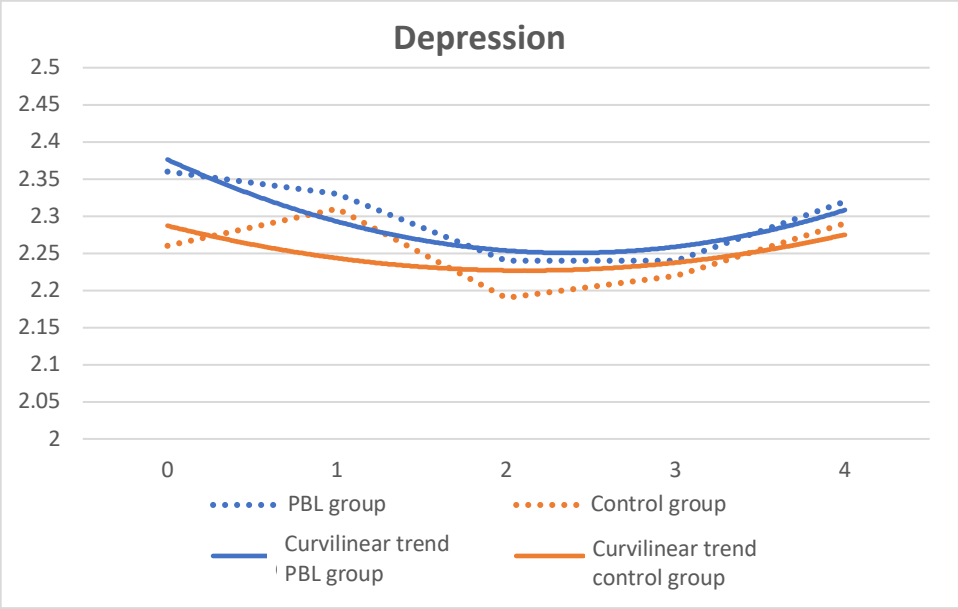
In addition, we conducted additional analysis in the schools that had PBL to examine whether applying PBL with greater fidelity also impacted on student outcomes. Results of general mixed models support the conclusion that conducting PBL with high fidelity will have a greater protective effect on students (including lower depression and generalised anxiety, greater positive affect, greater support and safety, greater school identification and greater school climate).

Overall, these results offer support for the positive impact of PBL on students (compared to no PBL), and its ability to promote school identification (a key mechanism in promoting positive outcomes for students), particularly when adopted with high fidelity.

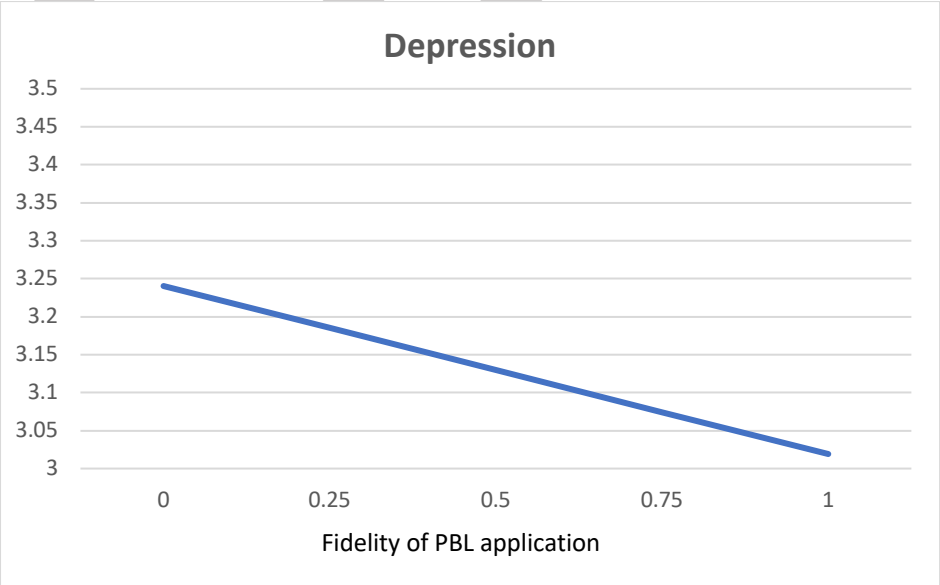
Mental Health

In this section we present the results for depression, generalised anxiety and positive affect.

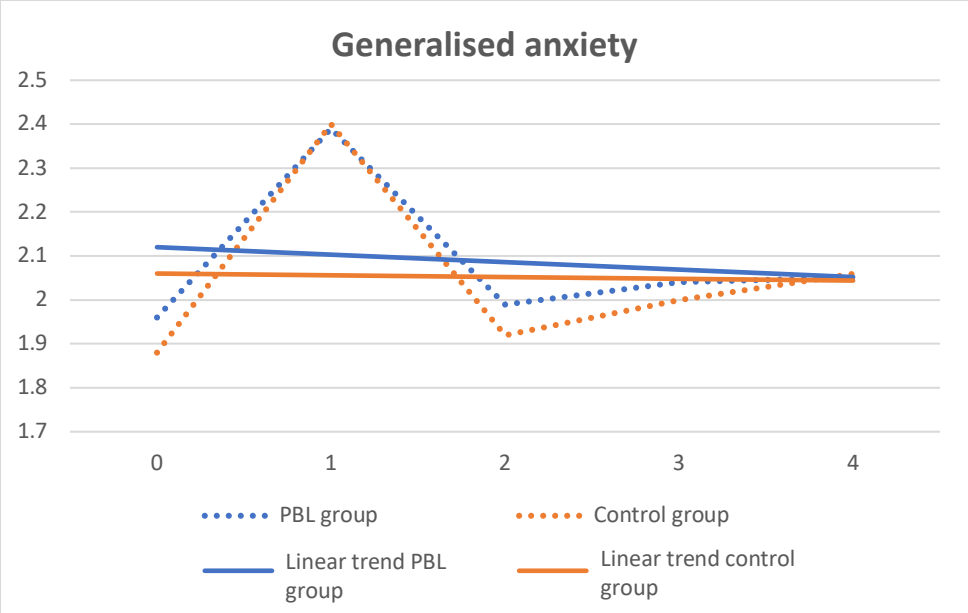
1. For depression, there was significant interaction between Year (or how much time had gone by) and Group (PBL versus control group; F [2.69, 1924.67] = 22.89**,** $p < .001$). This indicated that the evolution of depression across the five years was different for the PBL versus the control groups. Moreover, the results indicate that the evolution of depression in time was linear (or straight, F [1, 6668] = 8.78, $p < .001$) indicating a decrease in time, but a curved effect was also significant (F [1, 6668] = 29.24, $p < .001$) indicating a bent or change in direction. As illustrated in the figure below, student depression scores decreased more rapidly and steadily in PBL schools compared to control schools until Year 2.



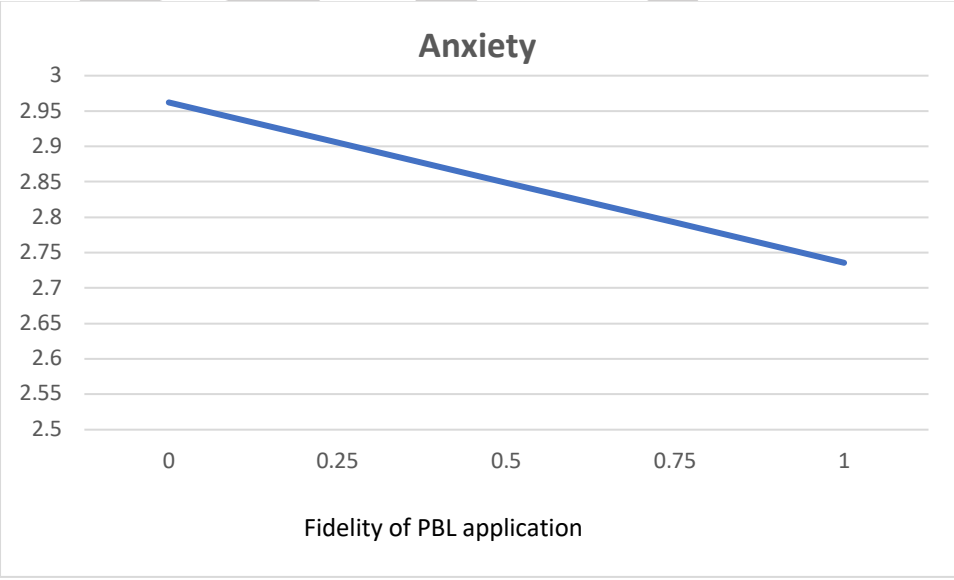
We also examined the effect of fidelity of PBL application in schools that had engaged in PBL. Results show that applying PBL with greater fidelity was associated with lower depression (b = -0.22, SE = 0.09, p = .024).



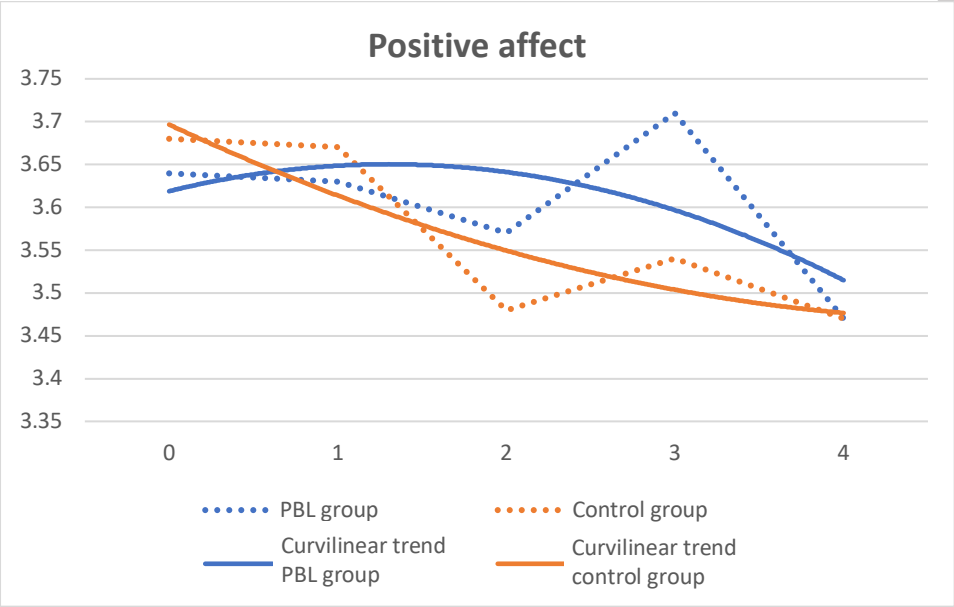
2. For generalised anxiety, there was significant interaction between Years and Group (F [3.27, 21863.50] = 15.68**,** $p < .001$). This indicated that the evolution of generalised anxiety across the five years was different for the PBL versus the control groups. Moreover, the results indicate that the evolution of generalised anxiety in time is linear (or straight, not curved; F [1, 6668] = 11.30**,** $p < .001$). As illustrated in the figure below, PBL started with somewhat higher generalised anxiety than the control group and at the five-year point groups were more similar with respect to levels of generalised anxiety suggesting the intervention in PBL facilitated the decrease in generalised anxiety compared to the control group.



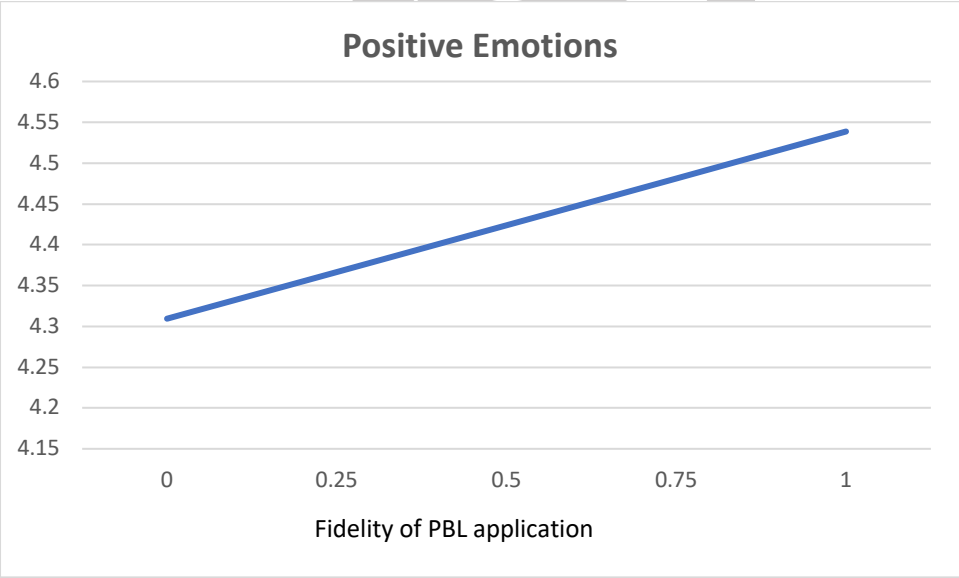
Analysis of the effect of PBL fidelity shows that PBL fidelity predicted lower generalised anxiety (b = -0.23, SE = 0.09, p = .016).



3. For positive affect, there was significant interaction between years and group (F [3.18, 21183.35] = 35.40, $p < .001$). This indicated that the evolution of positive affect across the five years was different for the PBL versus the control groups. The results also indicate that positive affect changes in a linear (F [1, 6668] = 28.45, $p < .001$) and quadratic fashion in time (F (1, 6668) = 30.08, $p < .001$). As illustrated in the figure below, PBL schools started lower and finished with higher positive affect than the control group suggesting the positive effect of the intervention in increasing positive affect until Year 2 and preventing less decrease in positive affect from Year 3.



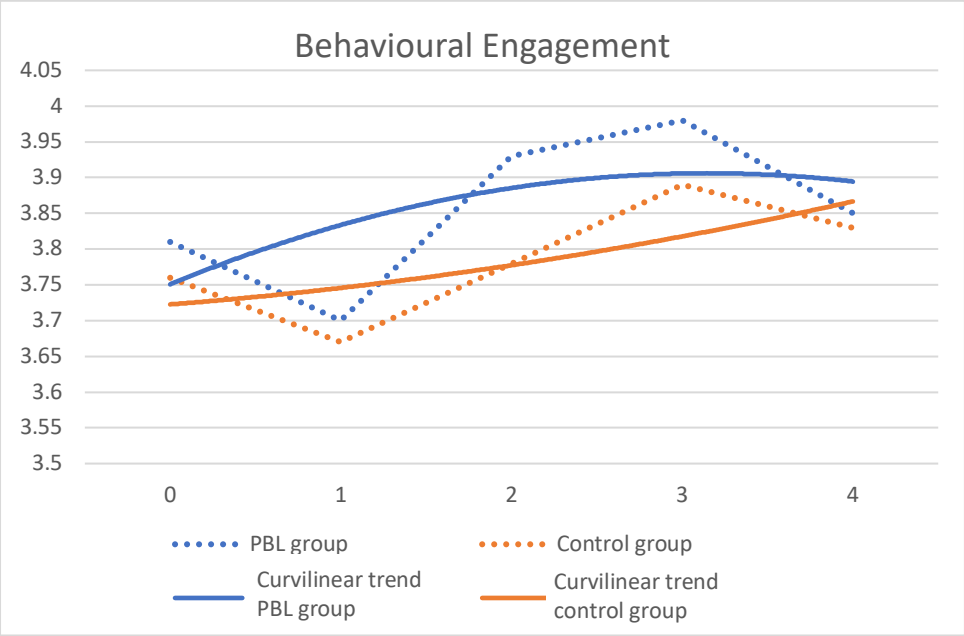
In terms of PBL fidelity, results show that greater PBL fidelity was not accompanied by greater positive affect. ($b = 0.23$, $SE = 0.19$, $p = .245$).



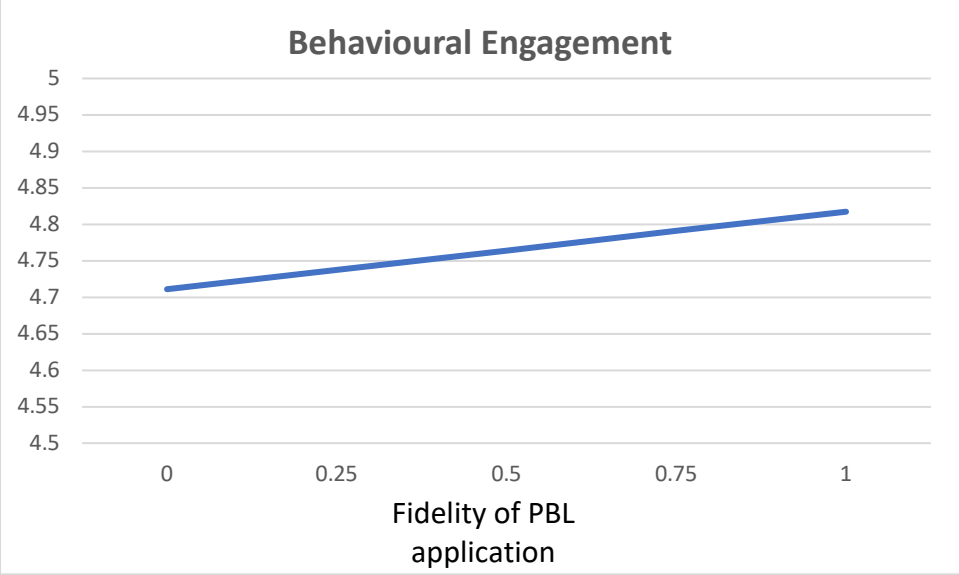
Learning Engagement

In this section we present the results for behavioural engagement, emotional engagement, and support and safety.

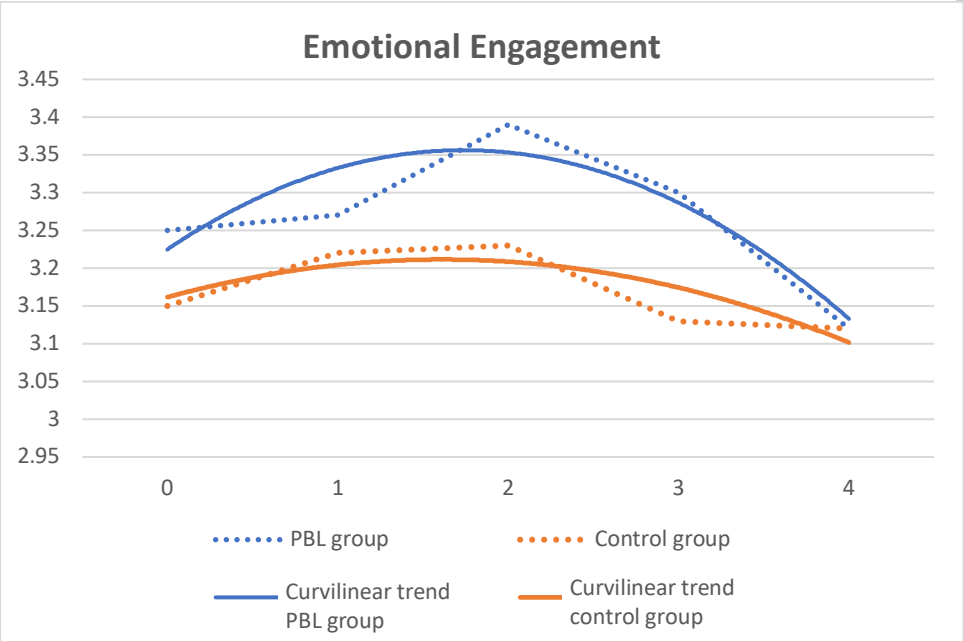
1. Concerning behavioural engagement, there was significant interaction between Years and Group (F [3.95, 18103.20] = 20.43**,** p < .001). This indicated that the evolution of behavioural engagement across the five years was different for the PBL versus the control groups. In addition, the results provide evidence for a quadratic evolution of behavioural engagement in time (F [1, 6668] = 49.66**,** p < .001; but not linear; F [1, 6668] =0.07 p = .792). As illustrated in the figure below, PBL groups evolved similarly to the control group in the early years but later showed a greater improvement in behavioural engagement.



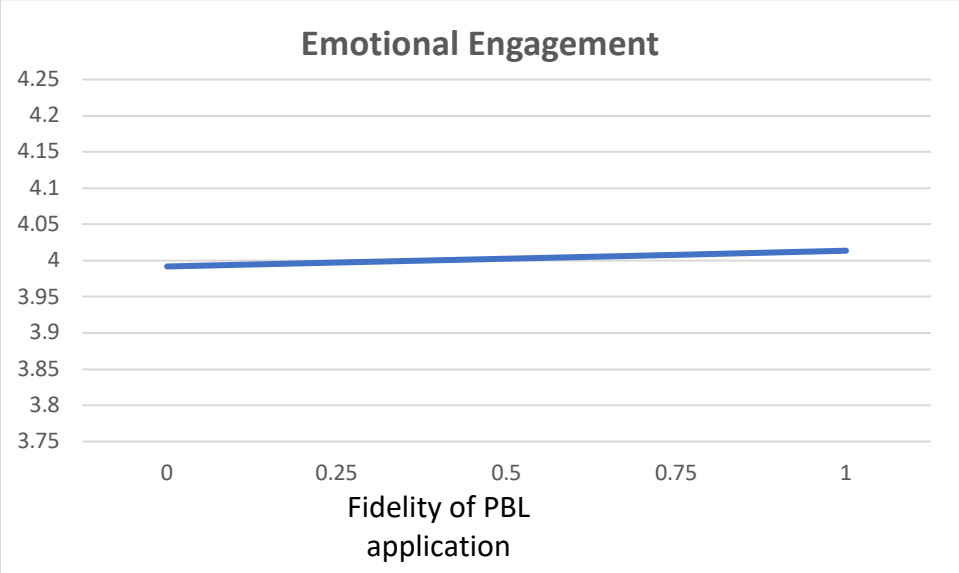
In contrast to other outcomes, applying PBL with greater fidelity was not significantly associated with behavioural engagement (b = 0.11, SE = 0.11, p = .952).



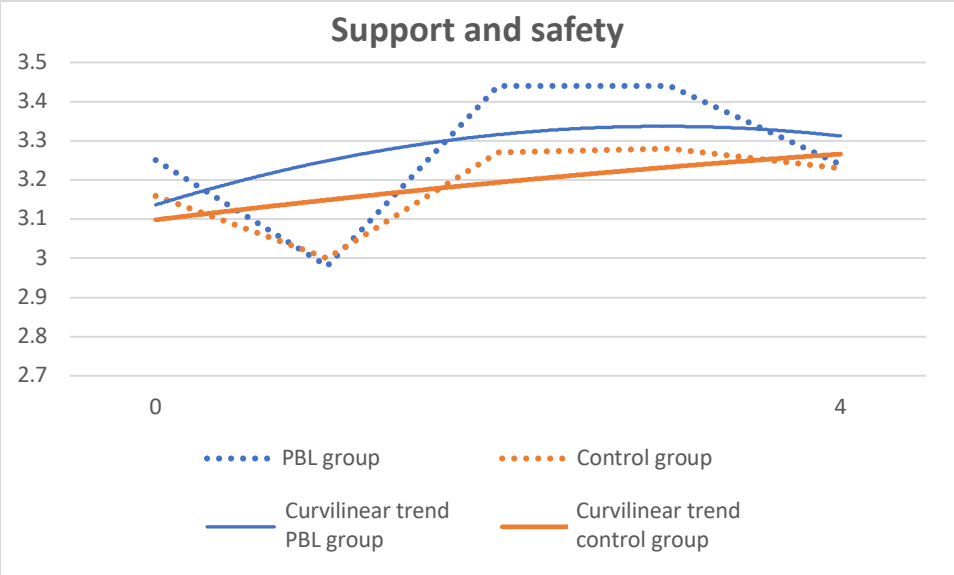
2. For emotional engagement, there was significant interaction between Years and Group (F [3.23, 21564.55] = 26.53**,** p < .001). This indicated that the evolution of emotional engagement across the five years was different for the PBL versus the control groups. Similarly to behavioural engagement, emotional engagement also showed evidence of a quadratic evolution in time (F (1, 6668) = 18.18**,** p < .001; but not linear; F [1, 6888] = 0.73**,** p = .393). As the figure below shows, PBL groups showed greater improvement in emotional engagement compared to the control group, with a drop off in the last year such that the groups were more similar.



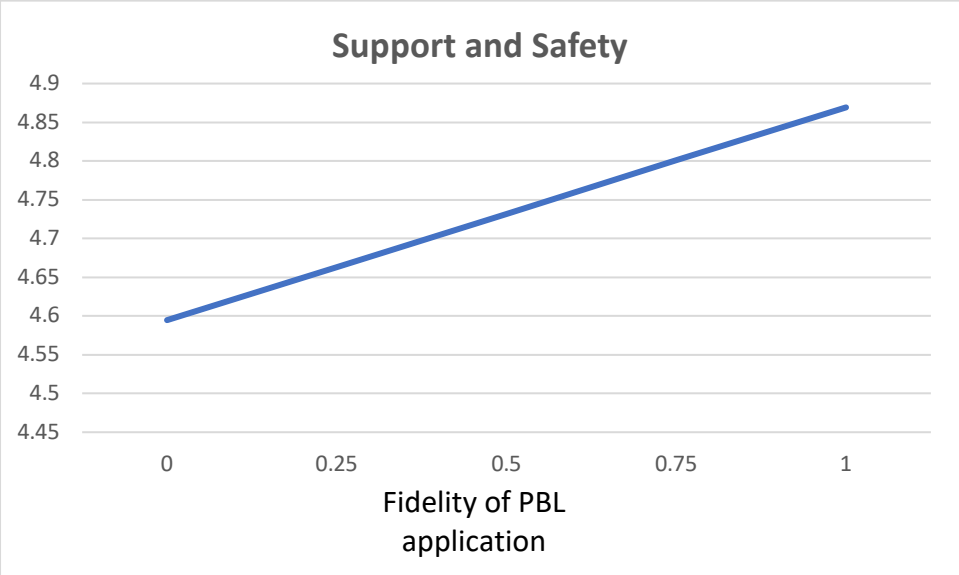
Here again, fidelity of PBL application was not significantly related to emotional engagement (b = 0.02, SE = 0.14, p = .872).



3. For support and safety, there was significant interaction between Years and Group (F (3.25, 21660.78) = 44.71, $p < .001$). This indicated that the evolution of support and safety across the five years was different for the PBL versus the control groups. Once more, support and safety showed evidence of a quadratic evolution in time (F (1, 6668) = 29.01, $p < .001$; but not linear, F (1, 6668) = 0.17, $p = .681$). As illustrated below, PBL groups showed greater improvement in support and safety compared to the control group, with a drop off in the last year such that the groups were more similar.



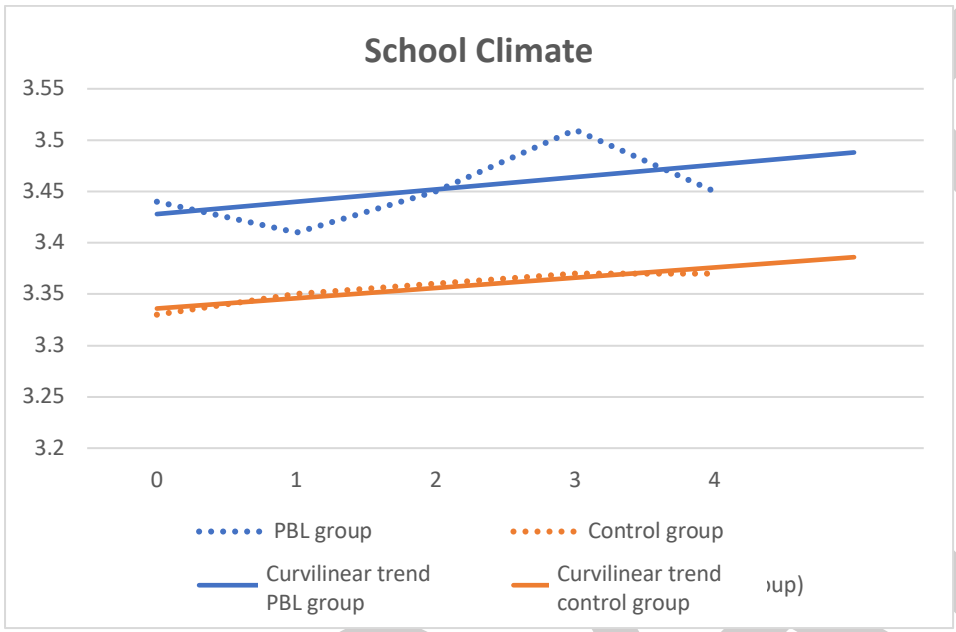
When considering the role of fidelity of PBL, fidelity of PBL application predicted greater perceived support and safety ($b = 0.27$, $SE = 0.13$, $p = .032$).



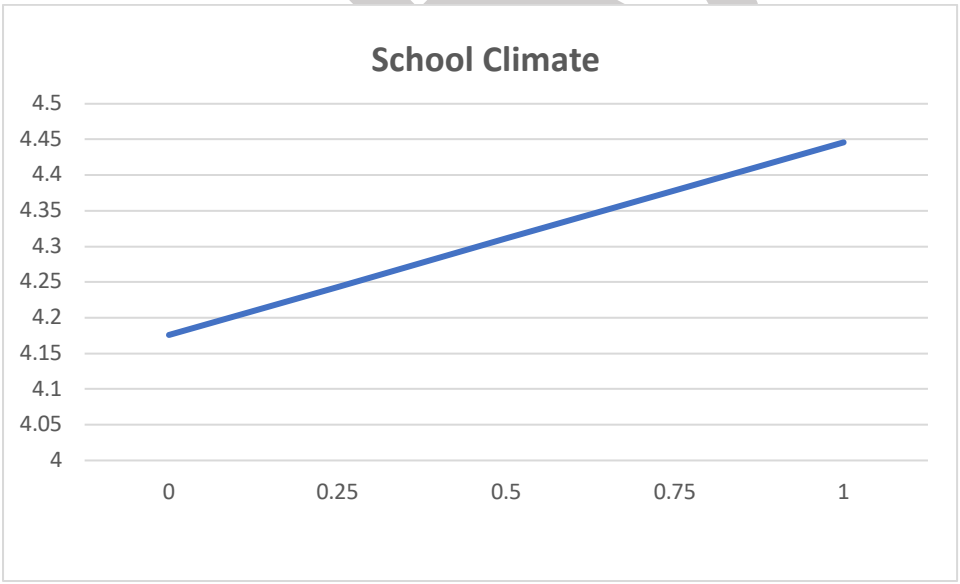
Perceptions of School Climate and School Identification

In this section we present the results for school climate and school identification.

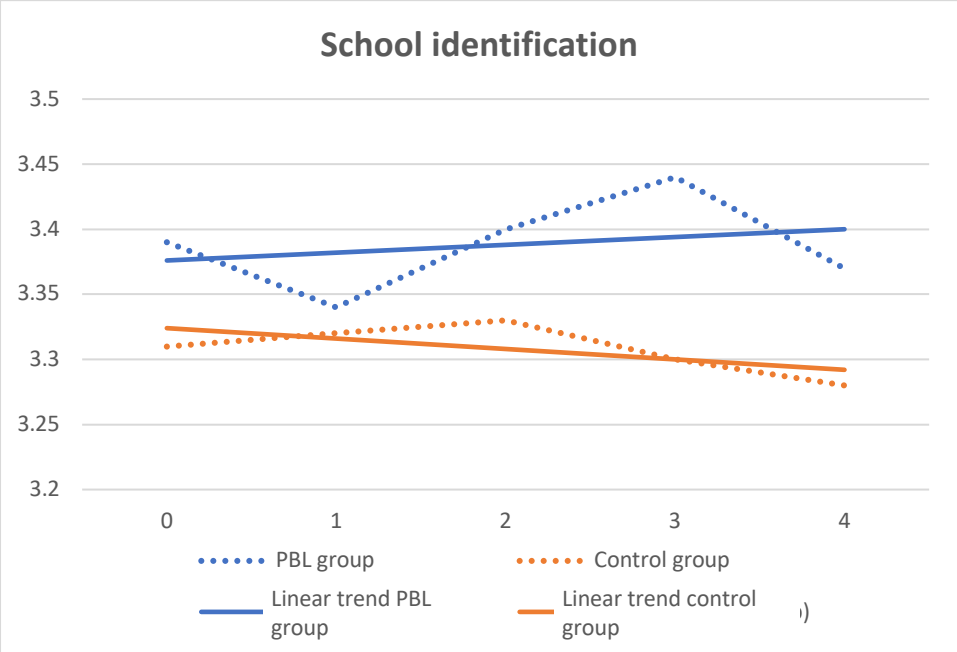
1. For school climate, we see that there was no significant interaction between Years and Group (F (3.93, 26472.36) = 1.51, p = .198). This indicates that school climate evolved in similar ways for both the PBL and control groups. As seen in the figure below, the PBL group started with greater perceived school climate and remained at similar levels compared to the control group.



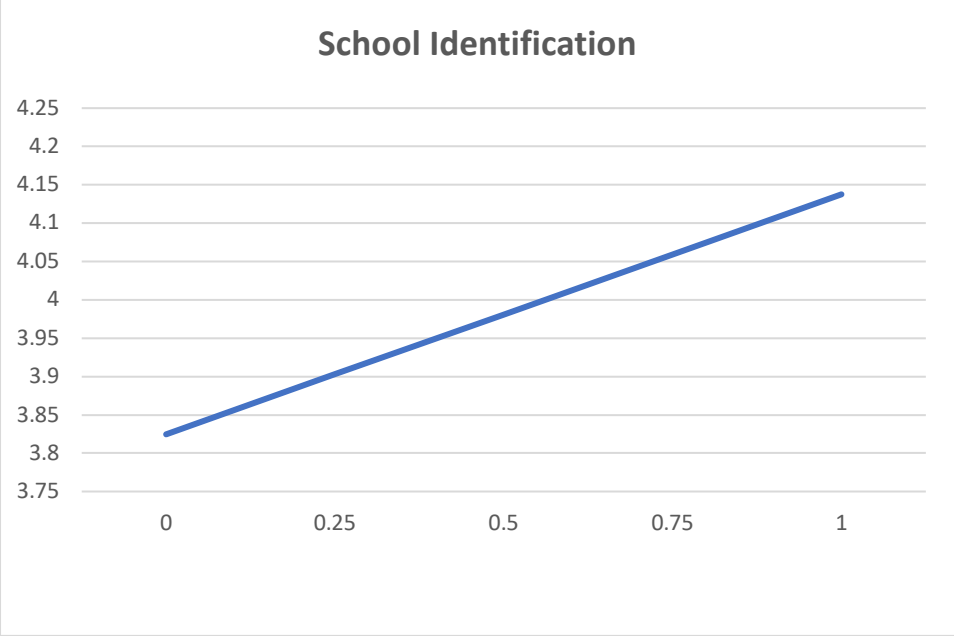
The analyses of fidelity of PBL showed that schools who implemented PBL with greater fidelity also showed increased school climate (b = 0.27, SE = 0.10, p = .005).



2. For school identification, there was a significant interaction Years and Group (F (3.93, 26300.14) = 1.18**,** p = .317). In other words, identification with school evolved differently in PBL versus the control group. School identification showed evidence of a linear evolution in time (F (1, 3198) = 6.50**,** p = .011). As the results below illustrate, the control group saw a decline of school identification in time, trended upward in time (a linear relation).



PBL fidelity was also related to greater school identification (b = 0.31, SE = 0.13, p = .015).



Discussion

How to Improve on PBL? Building a positive school climate and school identification

The results of this project have provided support for PBL as a school intervention capable of protecting students and improving student outcomes with some caveats concerning its long-term effects (given the curvilinear relation). This was demonstrated with a longitudinal and rigorous (data independent from PBL teams) study design, indicating that the empirical test of the positive effects of PBL reaches a high research standard. The strongest PBL intervention effect was suggested for the behavioural engagement outcome which is in line with the PBL focus. In addition, the fidelity effects were significant in explaining almost all outcomes, which presents confident evidence of the PBL intervention when applied as prescribed.

However, important improvements can be made. Specifically, we observe that PBL did increase students' school identification but not their perceptions of school climate. This means that the current effect of PBL can be amplified so that it may improve school climate and in turn school identification, mental health and student engagement. Specifically, PBL would be most effective at increasing school identification and student outcomes if it is capable of creating a positive school climate. In line with this, we present four recommendations to improve the effectiveness of Tier 1 PBL framework.

First, the PBL team decides and teaches expected behaviours. These expected behaviours are usually 3 to 5 positively-stated and easy to remember statements (e.g., be respectful). To be most effective, behavioural expectations need to go beyond defining behaviour and extend into actually defining what the schools stand for, i.e., their school climate. Behavioural expectations need to clarify and be intrinsically tied to who schools are, what they do, and how they do it. For example, the behavioural expectation of "being respectful" should become "we at [Name of] High school are respectful". Tying behavioural expectations to the schools' climate will result in greater compliance with these expectations among students who already identify with the school. It will also allow students who do not identify strongly with the school but do engage in the expected behaviours to develop a stronger sense of identification with the school.

Second, to be most effective in clarifying and improving school climate, PBL teams should consider the process by which they establish behavioural expectations, and particularly the inclusion of the student body in the process. There are three specific benefits to engaging the student body when selecting the expected behaviours. 1) If the student body (and staff) is involved in the creations of these expectations, they are more likely to perceive these rules as fair, representative of students and/or that they are for their benefit (i.e., procedural justice); 2) This sense of procedural justice will promote compliance with rules and laws; And 3) it will make clear for students what the school climate is, that it is well valued, and they as students can contribute to define the school climate (its values and approach, and the positive relations between students and between students and staff).

Third, the PBL intervention is expected to address the specific needs of the specific schools. Therefore, using currently available high-quality longitudinal data of each school is key to identifying weaknesses and strengths in school climate. To this end, it would be useful if PBL teams employ each schools' results on the School Climate and School Identification Measurement Scales-Student (SCASIM-St) to assess school needs. The results of SCASIM-St would allow PBL teams to clearly identify the aspects of school climate that are the weakest from students' perspectives and thus need to be reinforced via PBL intervention. In line with the core guiding principles of Tier 1

PBL (the use of data to make decision and monitoring student progress), the SCASIM-st data can also be used to examine students' perspective on school improvements/declines, a perspective that is often forgotten in current PBL research (and therefore implementation).

Finally, it is clear that fidelity of PBL application is important for better student outcomes. It is therefore essential for PBL to be applied according to the program development. This may be particularly hard at the beginning of the PBL process. Therefore, schools can be provided additional support during the first year so that they may apply a high-fidelity PBL program from the beginning and see faster improvements in students' lives.

Limitations

The data used in this report allowed us to match six ACT schools that had applied PBL since 2015/2016 with six ACT schools that had not applied PBL/had applied it later. This methodology enabled the comparison between these two groups of schools in a wide range of student outcomes. However, there are some limitations to the current analysis. To begin with, we are unaware of what is the response rate in our survey and therefore the extent to which our survey is representative of the student population. In addition, it would have been ideal to use hierarchical level (or multilevel model). This statistical model takes into consideration that students within a school have a lot in common because they are part of the same school. Comparing PBL to no PBL schools with multilevel modelling would have allowed us to distinguish the effects of PBL from different school properties. Unfortunately, multilevel modelling is considered appropriate with a minimum of 20 schools. Considering how we only had 12 schools in the current sample, this was insufficient for such complex modelling. In a similar vein, having five years of data offer important information about how outcomes change in time but it also means high rates of missing data. In the current report, a total number of 9952 students answered the survey at least one of the years, but the highest number of participants in any one year was 4060 (less than half of the total sample). We dealt with missing data by using multiple imputation when comparing the PBL group versus the control group, but other procedures (such as full information maximum likelihood) and other analyses (general linear model) that do not require multiple imputation could have been employed (as was done for the fidelity analyses). In addition, while the curvilinear effects informed the effect of PBL interventions, the decreasing trend mostly from Year 3 calls for future studies with longer-term modelling. Interestingly, the PBL curvilinear trend is opposite to the curvilinear developmental trend of student outcomes. The developmental trend shows that student outcomes deteriorate until grade 8 and then improving gradually. In our results, we see the opposite pattern: As a result of PBL, most outcomes improve until year 3 of the study, with this year 3 corresponding to critical grades (7^{th} 8^{th} and 9^{th} grades) for most students in our sample (n = 5667), i.e., the grades in which we find worst student outcomes. This suggests that PBL was most effective at improving outcomes as our current sample reached critical school grades Future research could examine whether PBL is indeed most useful in these critical grades.

While this report provides important information on student outcomes and perceptions, we could potentially examine other outcomes generally examined in PBL research. Specifically, most of the PBL research focuses on administrative outcomes (suspensions, major and minor referrals) and teachers perceptions of student's engagement and classroom behaviours. Given how we can potentially have access to this data, we could examine the effect of PBL on these various outcomes.

Conclusion

The research collaboration between the ACT Education Directorate and the Australian National University is well-positioned to answer a question puzzling education directorates worldwide: How can we improve students' well-being and engagement?

This report is an example of the high-quality research that this collaboration can produce. In this report, we outline for the first time whether PBL has an impact on students' self-reported well-being, engagement, school climate and school identification. Specifically, we provide evidence that PBL, an increasingly used school intervention/framework in the ACT, can improve student outcomes beyond what previous research has demonstrated with teachers' self-report and administrative data. This should provide further confidence to interveners, PBL teams and the Education Directorate on the value of PBL. There are two main implications to these findings that could be considered by the ACT Education Directorate.

1. PBL can be potentially more effective if it directly targets school climate.

2. PBL needs to be applied with high fidelity to be most effective.

Glossary

PBL: Positive behavioural learning is school-wide intervention/framework that aims at reducing problematic behaviours such as bullying and violence in school.

Generalised anxiety: A state of apprehension and psychic tension manifested as generalized worry. Example items measuring anxiety are: I worry about others liking me; I am nervous.

Depression: A condition of general emotional sadness and withdrawal that lasts for longer than is reasonable. Example items measuring depression are: I felt everything I did was an effort; I felt depressed.

Behavioural engagement: The extent to which students actively involve themselves in their own learning at school. Behaviourally engaged students would be expected to arrive to class on time, participate in activities, and apply themselves to the required tasks. Examples of items are: I try to complete my school work on time and to the best of my ability; I have got what it takes to be a good student.

Emotional engagement: How much students are interested in and enjoying learning at school. Emotionally engaged students would find the lessons fun and exciting, subsequently learning more. Examples of items are: I learn a lot from classes; I enjoy the work I do in class.

Fidelity: the extent to which school personnel are applying the core features of PBL. Core components are: defining the expectations, teaching these expectations, having a system that rewards students who meet expectations and a system that responds to behavioural violations, monitoring and decision making, management support, and the district-level support.

ICSEA: The index of community socio-educational advantage provides the mean level of socioeconomic advantage of the school.

Longitudinal design: A methodology in which data is gathered from the same people across several points in time.

Positive affect: The presence of positive emotions such as feeling rested and relaxed. Example items measuring positive affect are: Relaxed and free of tension; You generally enjoyed things.

Relational aggression: The presence of relational bullying, including rejection and malicious gossip. Example items measuring relational aggression are: I spread rumours/gossiped about others (e.g., at school, via SMS or on-line); I got angry at someone and deliberately ignored them/stopped talking to them

School climate: The social aspects of a school, which includes a sense of shared values and norms, an academic emphasis that supports learning, and positive relations between students and between students and staff members. Example items measuring school climate are: Students and staff are working towards the same goals; Teachers set high standards for learning in their classes; Students are friendly to each other.

School identification: Students' sense of connectedness or belonging in school; The extent to which being a member of the school is important to them. Example item are: Being a part of this school is important to me; I am happy to be a part of this school.

Support and safety: Students' perceptions that the school environment is safe and supportive of their own learning experience. Example items are: I feel close to others at this school; I feel safe at this school.

References

Allen, K.-A. (2019). Schools need to foster a sense of belonging. *The Sydney Morning Herald.* https://www.smh.com.au/education/schools-need-to-foster-a-sense-of-belonging-20190215-p50y16.html

Bizumic, B., Reynolds, K. J., Turner, J. C., Bromhead, D. & Subasic, E. (2009). The role of the group in individual functioning: School identification and the psychological well-being of staff and students. *Applied Psychology: An International Review, 58,* 171-192. (*Special issue on social identity, health and well-being)*

Bradshaw, C. P., Koth, C. W., Bevans, K. B., Ialongo, N., & Leaf, P. J. (2008). The impact of school-wide positive behavioral interventions and supports (PBIS) on the organizational health of elementary schools. *School Psychology Quarterly*, *23*(4), 462. https://doi.10.1037/a0012883

Bradshaw, C. P., Koth, C. W., Thornton, L. A., & Leaf, P. J. (2009). Altering school climate through school-wide positive behavioral interventions and supports: Findings from a group-randomized effectiveness trial. *Prevention Science, 10*, 100–115. https://doi.org/10.1007/s11121-008-0114-9

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on child behavior problems. *Pediatrics*, *130*(5), e1136-e1145. https://doi.org/10.1542/peds.2012-0243

Childs, K. E., Kincaid, D., & George, H. P. (2016). The relationship between school-wide implementation of positive behavior intervention and supports and student discipline outcomes. *Journal of Positive Behavior Interventions, 18,* 89–99. https://doi.org/10.1177/1098300715590398

Estrapala, S., Rila, A., & Bruhn, A. L. (2020). A Systematic Review of Tier 1 PBIS Implementation in High Schools. *Journal of Positive Behavior Interventions*, 1098300720929684.

Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions*, *11*(3), 133-144. https://doi-org.virtual.anu.edu.au/10.1177/1098300709332067

Maxwell, S., Reynolds, K. J., Lee, E., Subasic, E., & Bromhead, B. (2017) The Impacts of School Climate Perception and School Identification on Academic Achievement: Multilevel Modelling with Student and Staff/Teacher Response. Frontiers in Psychology. https://doi.org/10.3389/fpsyg.2017.02069

Noltemeyer, A., Palmer, K., James, A. G., & Wiechman, S. (2019). School-wide positive behavioral interventions and supports (SWPBIS): A synthesis of existing research. *International Journal of School & Educational Psychology*, *7*(4), 253-262. https://doi.org/10.1080/21683603.2018.1425169

Turner, I., Reynolds, K. J., Lee, E., Subasic, E., & Bromhead, D. (2018). Understanding aggression and victimization: Negative binomial modelling with supportive school climate, mental health, and social identity mediation. Translational Issues in Psychological Science, 4, 380-402. http://dx.doi.org/10.1037/spq0000074

Waasdorp, T. E., Bradshaw, C. P., & Leaf, P. J. (2012). The impact of schoolwide positive behavioral interventions and supports on bullying and peer rejection: A randomized controlled effectiveness trial. *Archives of Pediatrics & Adolescent Medicine, 166* (2), 149-156. https://doi.org/10.1542/peds.2012-0243

Appendices

Appendix 1: Measures

Appendix 2: Detailed Results

Appendix 1: Methodological details

Procedure and Data

Data for this research was collected as part of an ongoing longitudinal project between the Australian National University (ANU) and the Australian Capital Territory (ACT) Education Directorate (ED) (Reynolds, Subasic, & Bromhead, 2012). One aim of the project is to further understand and improve student outcomes by focusing on the school social environment (i.e., school climate and school identification). In a district with a population of nearly 400,000 (ABS, 2017), the project involved all 86 public schools. To test the effect of PBL on school outcomes, we used:

1. students responses from 2015 to 2019 (except for two of the schools, one who started PBL one year earlier and its matched school; for these schools, we used 2014-2018 data). The sample size for each year of the project was as follow: $n_{\text{year 0}} = 2380$; $n_{\text{year 1}} = 2396$; $n_{\text{year 2}} = 2574$; $n_{\text{year 3}} = 3497$; $n_{\text{year 4}} = 4060$. The total sample size was of $N_{all\ years} = 9952$. In line with national guidelines, the ethics review committee did not require parental consent given the low risk nature of the research and students being deemed capable of offering consent (see Section 4.2.9 National Health and Medical Research Council Guidelines). Students provided their consent and then completed an online survey (via Qualtrics software);
2. Data from the education directorate records included parental education, staff retention rate, school ICSEA, student enrolment, and PBL fidelity scores.

In this report, we make use of a natural experiment that occurred in the ACT. In 2015, the ACT Education Directorate began implementing PBL in several elementary and high schools of the territory. Simultaneously, in entire independence of the PBL team, the ongoing project between ANU and the ACT Education Directorate assessed students' mental health, engagement and school climate as well as school identification longitudinally. Therefore, we can examine differences in student outcomes as a result of adopting PBL in the school or not, and this independent of the PBL team's assessments.

The longitudinal project between ANU and the ACT Education Directorate gathers data of ACT students from 5th to 12th grade. Therefore, most students in the project are in P-10 (5th to 10th grades) or high schools (7th to 10th grades). Therefore, to maximize the number of students in our sample per school, this report focuses on data from students in P-10 or high schools. We identified six P-10 or high schools that had adopted PBL in 2015 or 2016 and matched these schools to six schools that have never applied PBL or adapted it much recently. These schools were matched based on ICSEA and school size (see Table 1).

Participants

The response rates represent student absences on survey days and difficulties in matching participants across waves and/or to the administrative data. Participants in year 0 of this project (2014/2015; pre –PBL) reported a mean age of 12.69 (*SD* = 1.84), with a similar sex split (male = 5077; female = 4669; missing across all years = 206). We assigned students the sex they had reported most often. When two sexes had been reported across different years, one of the options was always the "other" option, followed by non-sensical options (e.g., a helicopter, an alien). Therefore, the normative sex (male/female) split was assumed (as no participant mentioned only "other" across the years).

To calculate students grade at year 0, we used the grade participants reported in year 0. For those participants who did not respond to the survey in year 0, we used the grades they provided in later years (year 1 to year 4) to calculate what their grade 0 would have been. For example, if a student reported in year 2 being in 8th grade, then they were assigned "6th grade" in year 0. In this way, we maximized the number of valid and non-missing values. Grade frequencies at year 0 were as follows: Grade 1 = 208; Grade 2 = 260; Grade 3 = 1206; Grade 4 = 1405; Grade 5 = 1315; Grade 6 = 1495; Grade 7 = 1452; Grade 8 = 1123; Grade 9 = 997; Grade 10 = 390; Grade 11 = 44; unknown =

57). Most participants reported having at least one parent with a university degree ($n_{\text{with university degree}}$ = 3880; $n_{\text{without university degree}}$ = 2883; n_{missing} = 3189).

Measures

Table 1: PBL and control schools

PBL/Control schools	School	Sector	Year Joined	Sample size (2015/2016)	Size (2015)	ICSEA differ between P control sch (2015)
PBL	Calwell High School	High School	2015	7-10 *n* = 113	372	65
Control	Belconnen	High School	2018	7-10=231	356	
PBL	Caroline Chisholm School	P-10	2016	7-10 *n* = 227	586	72
Control	Mount Stromlo	p-10	Not yet	7-10=258	639 (2017)	
PBL	Kingsford Smith School	P-10	2016	7-10 *n* = 56	808	19
Control	Melrose High	High School	2019	7-10=262	707	
PBL	Namadgi School	P-10	2016	7-10 *n* = 84	595	64
Control	Gold Creek	p-10	2018	7-10=297	997	
PBL	University of Canberra High School Kaleen	High School	2016	7-10=97	238	64
Control	Campbell High	High School	2019	7-10=369	719	
PBL	Wanniassa School	P-10	2016	7-10 *n* = 133	422	7
Control	Melba High school	High School	2019	7-10 *n* = 190	547	

For the reliability of each scale in each of the years, see Table 2 below.

Depression. The Centre for Epidemiological Depression Scale (CES-D) Boston short-form (10 items) was used to measure depression (Kohout et al., 1993). The CES-D was developed to measure current symptoms of depression and has demonstrated internal reliability and content validity when used in adolescent populations (Chabrol et al. 2002; Cuijpers et al., 2008; Herge et al., 2013) and includes items such as "I felt everything I did was an effort" and "I felt depressed". Students rated the items on a 5-point Likert scale that ranged from 0 (Rarely/none) to 4 (Very often/always). Items were averaged together such that higher scores suggest greater level of depressive symptomology.

Generalised anxiety: Nine items based on the Screen for Child Anxiety Related Emotional Disorders measure (SCARED; Birmaher et al., 1997) were used to measure generalised anxiety. SCARED was developed to screen children with anxiety, showing good internal consistency, discriminant validity, and test-rest reliability in previous research (Birmaher et al., 1999; Birmaher et al., 1997). The items include statements such as "I am nervous" and "I worry about being as good

as other kids". Students rated items on a 3-point Likert scale that ranged from 0 (Not true or hardly true) to 2 (True or often true). Items were averaged such that higher scores indicated more anxiety.

Positive affect. The personal well-being subscale of the Australian Adolescent version (Heybeck & Neill, 2000) of the Mental Health Inventory (MHI; Veit & Ware, 1983) was used to measure general positive affect (10 items). Students rated how often they experienced positive emotions (e.g., "Relaxed and free of tension" and "You generally enjoyed things") in the past month, with a 8-point scale ranging from 0 (None of the time) to 7 (All of the time). The MHI has demonstrated good internal consistency (Heubeck & Neill, 2000), content validity (Cassileth et al., 1984; Ware et al., 1979) and test-restest reliability (Heubeck & Neill, 2000; Veit & Ware, 1983). The 10 items were averaged, and higher scores indicate more positive affect.

Emotional Engagement. Four items were used to measures emotional engagement. These items were drawn from Wellborn (1991), Fredricks et al. (2003), and Skinner et al. (2013), and assess the extent to which individuals are interested in and enjoy learning at school (e.g., "I feel excited by my work at school"; "I am interested in what I learn at school"). The mean of items (answered using a Liker-scale ranging from 1 [Strongly disagree] to 5 [Strongly agree]) was calculated such that higher numbers indicated greater emotional engagement.

Behavioural engagement. Six items measured behavioural engagement, or the extent to which students' behaviour show engagement in learning outcomes. This measure was adapted from several scales from Bizumic et al. (2009), Skinner et al. (2013), and Caraway et al. (2013) and responded on a 5-point Likert scale ranging from 1 (Strongly disagree) to 5 (Strongly agree). Examples of items are: I actively participate in class activities; I try to learn at school as much as I can. The mean of items was calculated such that higher numbers of behavioural engagement indicate greater behavioural engagement.

Support and safety. Seven items were adopted from school research literature (e.g., Anderson-Butcher et al., 2012; McNeely, 2002; Brown & Evans, 2002) to assess the extent to which students feel safe (example of item: I feel safe at this school), supported (example of item: I feel I can talk to teachers about problems at school) and emotionally connected to their school (example of item: I enjoy coming to my school). Items were answered using a 5-point Likert scale ranging from 1 (Strongly disagree) to 5 (Strongly agree) and where averaged, with higher number indicates greater support and safety.

School climate and school identification. To measure school climate and school identification, a shortened-version of the School Climate and School Identification Measurement Scale-Student (SCASIM-St; Lee et al., 2017) was employed. The SCASIM-St used Moos' (1973) framework and social identity theory (Tajfel & Turner, 1979; Turner et al., 1987) to create a theoretically-driven scale. The SCASIM-St measures school climate with 26 items measuring the school climate dimension (its four sub-dimensions: academic emphasis [7 items], shared values and approach [6 items], student-student relations [6 items] and staff-student relations [7 items]) such that high scores reflect greater school climate. Similarly, the five school identification items were averaged, and higher scores reflect greater identification. Items were answered using a Likert-scale ranging from 1 (Totally disagree) to 7 (Totally agree) in 2014-2016, with the Likert-scale changing after 2016 from 1 (Totally disagree) to 5 (Totally agree). To homogenize the responses, answers from 2014-2016 were rescaled to range from 1 to 5.

Fidelity. Fidelity was assessed by the PBL team every year using a modified version of the School-wide Evaluation Test (or SET), a standardized and validated scale frequently used in PBL research to assess the fidelity of the intervention (Horner, 2004). It examines seven subdimensions of fidelity: 1. the extent to which schools defined expectations; 2. whether they taught these expectations; 3. whether they had a system that rewards students who meet expectations; 4. Whether they a system that responds to behavioural violations; 5. how they monitored and were engaged in decision making; 6. whether they had support from management; 7. And whether they had support at the district-level. Each of these subdimensions were rated using a scale from 0 to 1, with 1 indicating

perfect fidelity to the PBL instructions. Fidelity was assessed for years 2 to 4 (therefore we are missing fidelity scores before PBL and in the first year of PBL) and only in the six schools that used the PBL framework.

Covariates. We controlled for three characteristics of the school: schools' socio-economic status (ICSEA), the number of students enrolled from 2016 to 2019 (averaged across the years) and staff retention rate from 2016 to 2019 (averaged across the years). As for students' characteristics, we controlled for parental education (whether at least one of the parents had reported having a university degree), student sex, and student grade at year 0.

Table 2. Cronbach alpha (internal consistency) across years

	Depression	Anxiety	Positive Affect	Behavioural Engagement	Emotional Engagement	Safety and Support	School Climate	School Identification
Year 0	.83	.89	.95	.95	.95	.93	.97	.93
Year 1	.84	.91	.96	.94	.94	.92	.97	.94
Year 2	.71	.91	.95	.91	.92	.90	.96	.92
Year 3	.84	.91	.95	.92	.92	.89	.96	.93
Year 4	.84	.91	.95	.93	.92	.89	.96	.93

Appendix 2: Statistical plan and detailed results

Data cleaning, missing data analysis, and descriptive statistics were completed using SPSS Version 26, with multiple imputation (MI) conducted in Mplus Version 8.4.

Data from the different schools and years were combined into a single data set that included all data from year 0 to year 4. Missing Values Analysis (MVA) showed a missing rate ranging from 58.5% to 77.8% for depression, anxiety, positive affect, emotional engagement, behavioural engagement, student safety and support, school climate and school identification.

For the individual covariates, the missing rate ranged from 32% for parental education to 0.6% Grade at year 0 (with no missing for ICSEA at year 0, staff retention rate and student enrolment). Little's MCAR test was significant ($p < .01$), suggesting that the data was not missing completely at random (MCAR), and missingness was also related to other variables in the dataset (i.e., missing at random [MAR]; Newman, 2014; Tabachnick & Fidell, 2013). To retain as many cases as possible when conducting mixed ANOVAs (Newman, 2003), multiple imputation (MI) with Mplus was used to impute missing values all outcome variables across the five years (depression, anxiety, positive affect, emotional engagement, behavioural engagement, student safety and support, school climate and school identification). As recommended (Graham, Olchowski, & Gilreath, 2007), forty datasets were imputed and then merged together to conduct the Mixed ANOVAs (but not the mixed effects models) in SPSS. Means and standard deviations after multiple imputation are presented in Table 3.

Mixed ANOVAs were conducted to test whether students in PBL schools differ from students in the control group (i.e., a categorical variable with control group coded as 0 and PBL group coded as 1) in the way their outcomes evolved in time (time being a continuous variable). Specifically, a mixed ANOVA tests whether an outcome (e.g., depression) changes across the years (i.e., the main effect of Years), whether the PBL and Control groups are, on average, different from each other in the outcome (i.e., the main effect of Groups) and the interaction between Years and Group (Years X Control) on a student outcome (e.g., depression). Critically, if this interaction is significant, it indicates that the way depression changes in time is different in the PBL versus control group. Graphs and the linear trends per group are then examined to understand this difference in evolution in time. Greenhouse-Geiser correction was used when testing the effect of Time and the interaction between GroupXTime given how Muachly's test showed that the sphericity assumption was not met for all outcomes.

To examine the effect of fidelity on student outcomes, we employed mixed effects modelling, which allows us to examine the longitudinal effect of a continuous variable (fidelity) on student outcomes. Since we have fidelity scores for three of the five years of the project, mixed models allow us to test the longitudinal variation of fidelity and its effect on student outcomes across time. In addition, mixed modelling allows us to account for each individuals' specificities (e.g., their own level of depression and their own rate of change), therefore accounting for greater variance while also allowing data to be incomplete or missing. Therefore, the imputed data set was not used for these analyses.

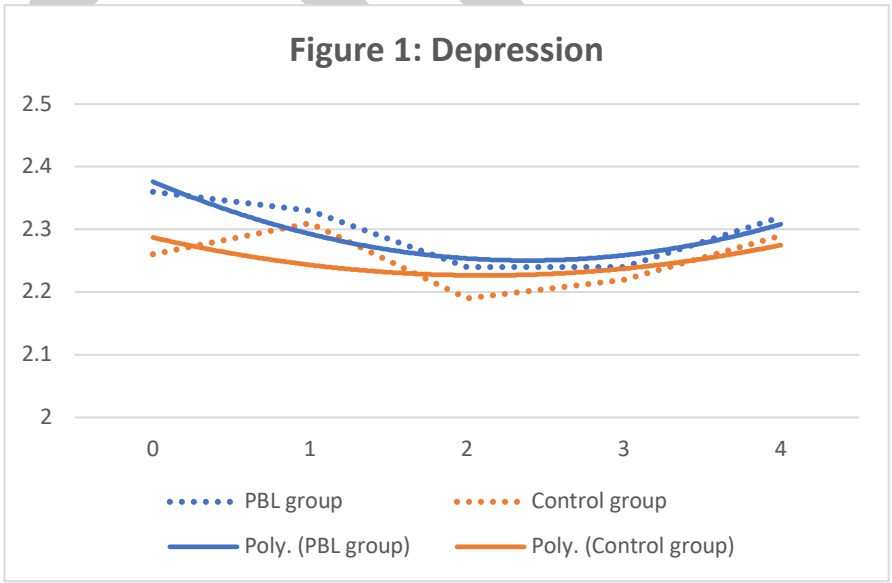
Table 3. Means and standard deviations across years									
	Depression *M* (*SD*)	**Anxiety *M* (*SD*)**	**Positive Affect *M* (*SD*)**	**Behavioural Engagement *M* (*SD*)**	**Emotional Engagement *M* (*SD*)**	**Safety and Support *M* (*SD*)**	**School Climate *M* (*SD*)**	**School Identification *M* (*SD*)**	**Fidelity *M* (*SD*)**
Year 0	2.29 (0.34)	1.90 (0.36)	3.66 (0.75)	3.76 (0.57)	3.18 (0.71)	3.18 (0.66)	3.35 (0.74)	3.31 (0.77)	-
Year1	2.31 (0.30)	2.38 (0.30)	3.63 (0.75)	3.66 (0.58)	3.23 (0.70)	2.98 (0.49)	3.37 (0.72)	3.32 (0.76)	-
Year 2	2.22 (0.28)	1.94 (0.39)	3.50 (0.78)	3.81 (0.54)	3.31 (0.66)	3.31 (0.64)	3.38 (0.69)	3.35 (0.72)	.49 (.21)
Year 3	2.24 (0.44)	2.02 (0.41)	3.59 (0.82)	3.90 (0.59)	3.24 (0.69)	3.33 (0.67)	3.41 (0.69)	3.35 (0.77)	.76 (.14)
Year 4	2.31 (0.45)	2.06 (0.41)	3.46 (0.86)	3.83 (0.64)	3.15 (0.74)	3.24 (0.70)	3.40 (0.69)	3.32 (0.80)	. 76 (12)
M = Mean; *SD* = Standard deviation									

Comparing PBL to control groups

Depression. To test whether students in schools that had PBL experienced a greater decrease in depression compared to students from the control group, a 5 (Time) X 2 (Group) mixed ANOVA was conducted with Time (year 0 to year 4) as the within-group factor and Group (control vs PBL) as the between-groups factor while controlling for ICSEA, sex, staff retention rate, student enrollment, grade, and parental education. We see in Table 4 that there was a main effect of both time (indicating that depression changed in time) and group (indicating that there were differences in the level of depression between the PBL and control groups). Importantly, the interaction between time and group was significant, which indicates that the groups differed in the evolution of depression in time. The evolution of depression across time in the PBL group was both linear (or straight, F [1, 6668] = 72.96, $p < .001$) and curved effect (F [1, 6668] = 15.93, $p < .001$), indicating a decrease in time followed by a turn or bend. A similar pattern emerges with the control group (linear trend, F [1, 6668] = 94.79, $p < .001$; quadratic trend, F [1, 6668] = 13.98, $p < .001$). As illustrated in Figure 1, depression decreased in both groups and then stabilized, with a greater decrease in depression in the PBL group.

Table 4. Results for depression

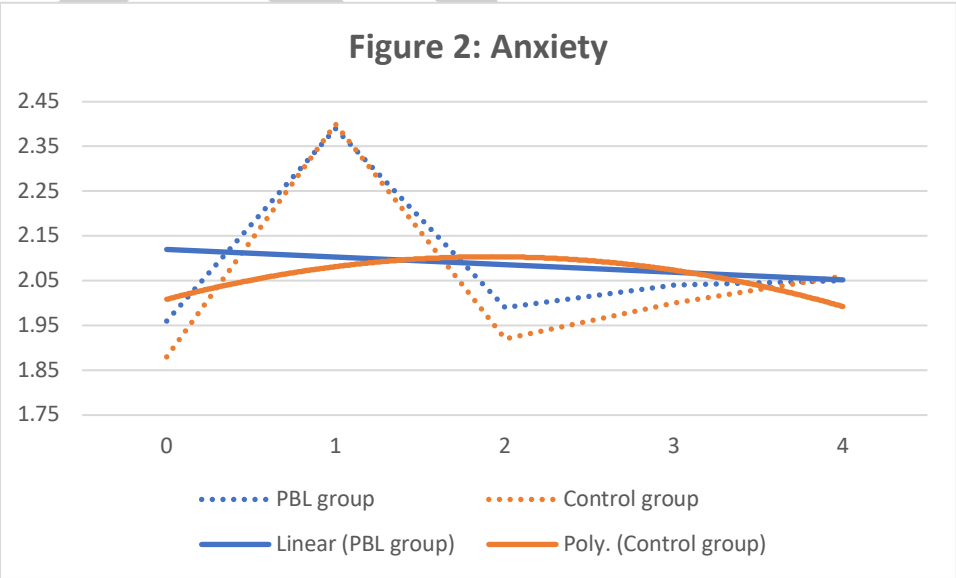
Predictor	Degrees of freedom	Mean Square	*F*	*p*
Intercept	1	82.07	200.77	<.001
Group (Control = 0; PBL = 1)	1	14.15	34.61	<.001
Time (Year 0 to Year 4)	2.69	9.68	90.36	<.001
Group X Time	**2.69**	**2.45**	**22.89**	**<.001**
ICSEA	1	0.69	1.70	.193
Sex	1	47.66	116.60	<.001
Staff retention rate	1	0.62	1.52	.217
Student enrollment	1	2.49	6.09	.014
Grade	1	5.78	14.12	<.001
Parental Education	1	1.02	2.51	.114
Error	6668	0.41		



Generalized anxiety. To test whether students in schools that had PBL experienced a greater decrease in generalized anxiety compared to students from the control group, a 5 (Time) X 2 (Group) mixed ANOVA was conducted with Time (year 0 to year 4) as the within-group factor and Group (control vs PBL) as the between-groups factor while controlling for ICSEA, sex, staff retention rate, student enrollment, grade, and parental education. We see that there was a main effect of both time (indicating that generalized anxiety changed in time) and group (indicating that there were differences in the level of generalized anxiety between the PBL and control groups). Importantly, the interaction between time and group was significant, which indicates that the groups differed in the evolution of generalized anxiety in time. The PBL group changed in linear fashion, decreasing significantly in time as seen in Figure 2 (F [1, 6668] = 5.58, p = .018 but not in quadratic fashion, F [1, 6668] = 0.51, p = .476). In contrast, the control group had only a quadratic change in time (F [1, 6668] = 4.96, p = .026 but no linear time, F [1, 6668] = 2.00, p = .158), with an increase in anxiety in time followed by a decrease.

Table 5. Results for anxiety

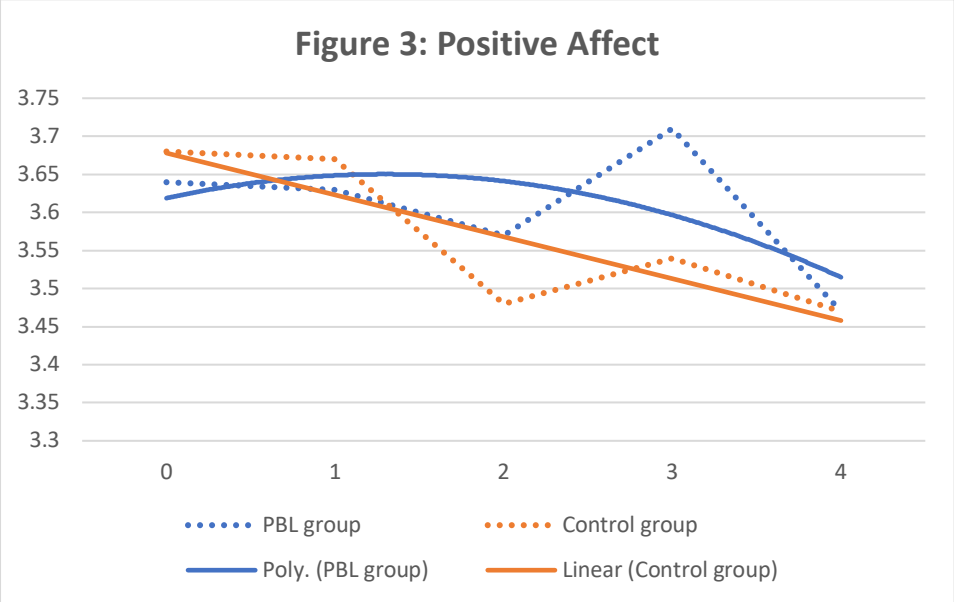
Predictor	Degrees of freedom	Mean Square	*F*	*p*
Intercept	3.28	1.21	17.10	<.001
Group (Control = 0; PBL = 1)	1	7.69	16.55	<.001
Time (Year 0 to Year 4)	2.69	9.68	90.36	<.001
Group X Time	**3.28**	**1.11**	**15.68**	**<.001**
ICSEA	1	0.12	0.27	.605
Sex	1	281.63	605.68	<.001
Staff retention rate	1	0.36	0.77	.382
Student enrollment	1	4.30	9.25	.002
Grade	1	18.87	40.58	<.001
Parental Education	1	0.57	1.23	.268
Error	6668	0.47		



Positive affect. To test whether students in schools that had PBL experienced a greater increase in positive affect compared to students from the control group, a 5 (Time) X 2 (Group) mixed ANOVA was conducted with Time (year 0 to year 4) as the within-group factor and Group (control vs PBL) as the between-groups factor while controlling for ICSEA, sex, staff retention rate, student enrollment, grade, and parental education. We see from Table 6 that there was no significant main effect of group (indicating that on average the control and PBL groups did not differ significantly) and a significant main effect of time (indicating that positive affect changed in time). Importantly, the interaction between time and group was significant, which indicates that the groups differed in the evolution of positive affect in time. The PBL group changed in quadratic fashion as seen in Figure 3 (F [1, 6668] = 16.13 $p < .001$ but not in linear trend, F [1, 6668] = 1.87, $p = .172$), starting lower in positive affect but then maintaining a higher level of positive affect that the control group. The control group showed a linear decrease in time (F [1, 6668] = 26.57, $p < .001$ but no quadratic trend, F [1, 6668] = 2.37, $p = .123$).

Table 6. Results for positive affect

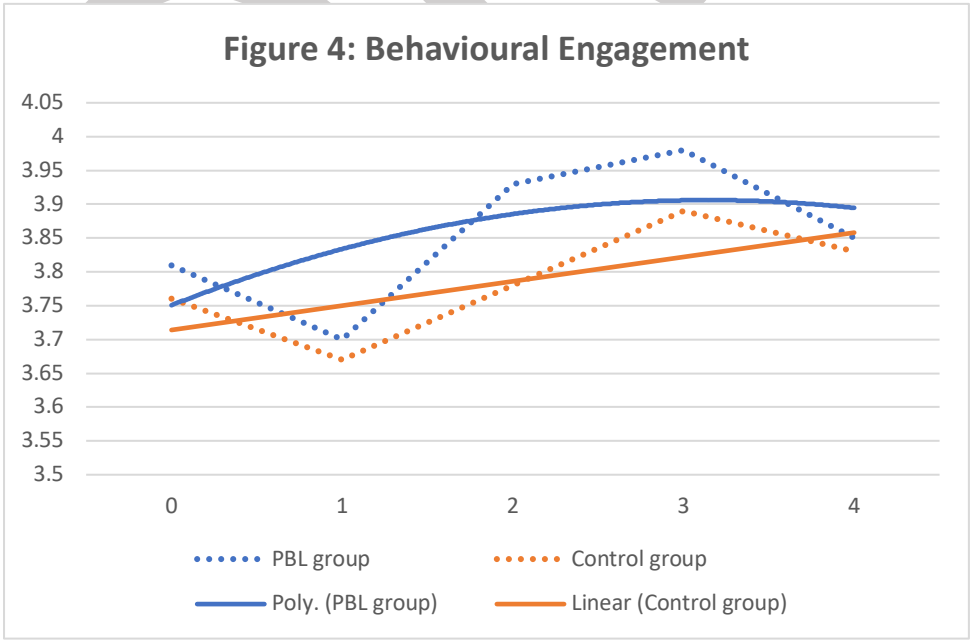
Predictor	Degrees of freedom	Mean Square	*F*	*p*
Intercept	1	29.24	12.84	<.001
Group (Control = 0; PBL = 1)	1	4.97	2.18	.140
Time (Year 0 to Year 4)	3.18	14.43	56.31	<.001
Group X Time	**3.18**	**9.07**	**35.40**	**<.001**
ICSEA	1	91.03	39.96	<.001
Sex	1	177.03	77.71	<.001
Staff retention rate	1	46.17	20.27	<.001
Student enrollment	1	104.69	45.96	<.001
Grade	1	92.11	40.43	<.001
Parental Education	1	13.69	6.01	.014
Error	6668	2.28		



Behavioural engagement. To test whether students in schools that had PBL experienced a greater increase in behavioural engagement compared to students from the control group, a 5 (Time) X 2 (Group) mixed ANOVA was conducted with Time (year 0 to year 4) as the within-group factor and Group (control vs PBL) as the between-groups factor while controlling for ICSEA, sex, staff retention rate, student enrollment, grade, and parental education. We see from Table 7 that there was a main effect of both time (indicating that behavioural engagement changed in time) and group (indicating that there were differences in the level of behavioural engagement between the PBL and control groups). Importantly, the interaction between time and group was significant, which indicates that the groups differed in the evolution of behavioural engagement in time. The PBL group changed in quadratic fashion as seen in Figure 4 (F [1, 6668] = 11.32 p = .001 but not in linear trend, F [1, 6668] = 1.66, p = .197), showing a general increase in time that becomes stabilized. In contrast, the control group did not show any significant change in time (linear, F [1, 6668] = 0.51, p = .473; quadratic, F [1, 6668] = 0.27, p = .601).

Table 7. Results for behavioural engagement

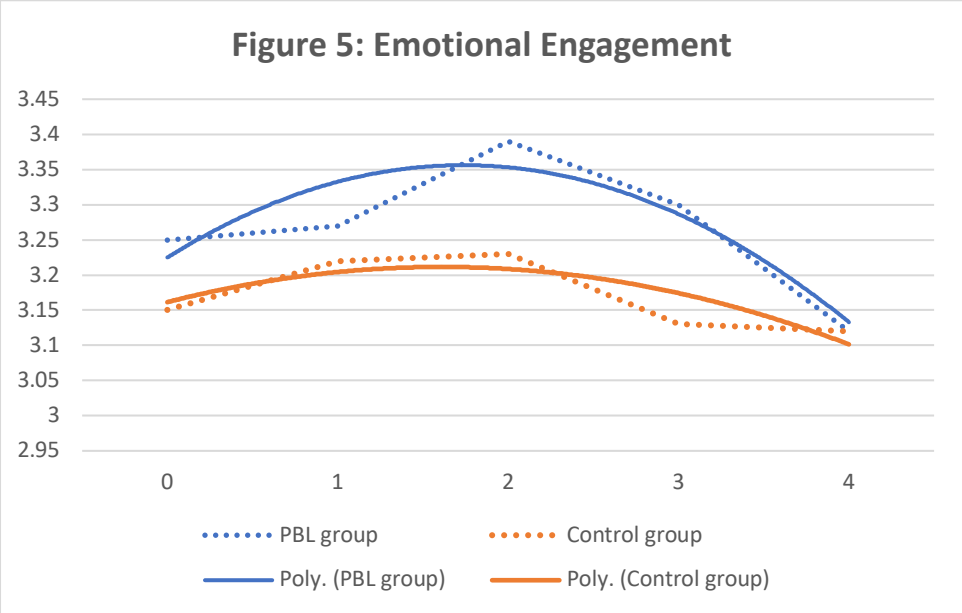
Predictor	Degrees of freedom	Mean Square	*F*	*p*
Intercept	1	143.95	128.30	<.001
Group (Control = 0; PBL = 1)	1	20.97	18.69	<.001
Time (Year 0 to Year 4)	2.72	3.76	19.48	<.001
Group X Time	**2.72**	**3.95**	**20.43**	**<.001**
ICSEA	1	16.03	14.28	<.001
Sex	1	39.02	34.78	<.001
Staff retention rate	1	0.04	.03	.856
Student enrollment	1	8.37	7.46	.006
Grade	1	122.99	109.61	<.001
Parental Education	1	34.74	30.96	<.001
Error	6668	1.12		



Emotional engagement. To test whether students in schools that had PBL experienced a greater increase in emotional engagement compared to students from the control group, a 5 (Time) X 2 (Group) mixed ANOVA was conducted with Time (year 0 to year 4) as the within-group factor and Group (control vs PBL) as the between-groups factor while controlling for ICSEA, sex, staff retention rate, student enrollment, grade, and parental education. We see from Table 8 that there was a main effect of both time (indicating that emotional engagement changed in time) and group (indicating that there were differences in the level of emotional engagement between the PBL and control groups). Additionally, the interaction between time and group was significant, which indicates that the groups differed in the evolution of emotional engagement in time. The PBL group changed in linear (F [1, 6668] = 17.86, $p < .001$) and quadratic fashion (F [1, 6668] = 5.40 $p = .020$), as seen in Figure 5 showing a greater increase in time compared to the control group (linear, F [1, 6668] = 11.33, $p = .001$; quadratic, F [1, 6668] = 4.82, $p = .028$), with a drop off in the last year such that the groups were more similar.

Table 8. Results for emotional engagement

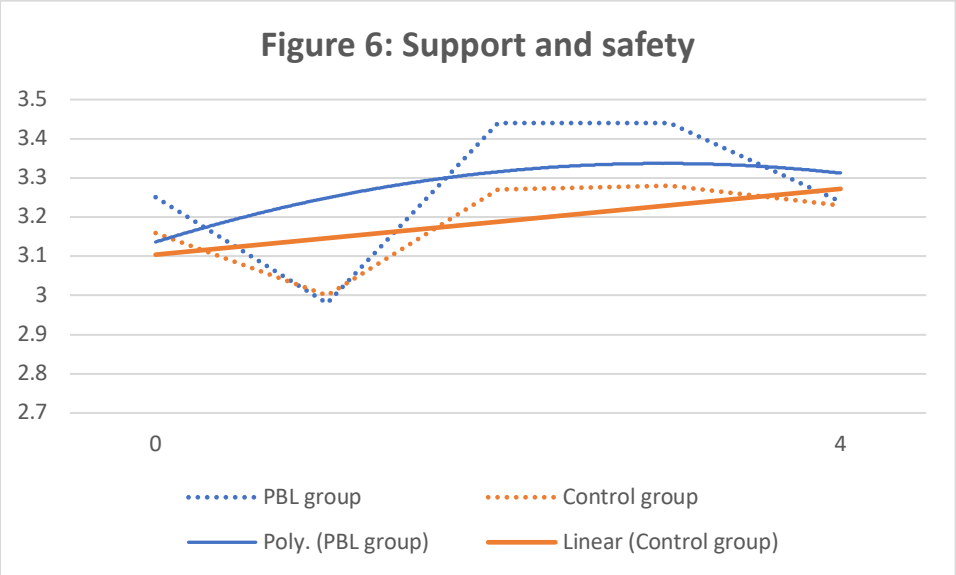
Predictor	Degrees of freedom	Mean Square	*F*	*p*
Intercept	1	95.44	52.80	<.001
Group (Control = 0; PBL = 1)	1	48.39	26.77	<.001
Time (Year 0 to Year 4)	3.23	4.31	22.81	<.001
Group X Time	**3.23**	**5.01**	**26.53**	**<.001**
ICSEA	1	9.99	5.53	.019
Sex	1	2.15	1.19	.275
Staff retention rate	1	4.20	2.33	.127
Student enrollment	1	.05	.03	.870
Grade	1	75.26	41.64	<.001
Parental Education	1	34.74	30.96	<.001
Error	6668	1.12		



Support and safety. To test whether students in schools that had PBL experienced an increase in support and safety compared to students from the control group, a 5 (Time) X 2 (Group) mixed ANOVA was conducted with Time (year 0 to year 4) as the within-group factor and Group (control vs PBL) as the between-groups factor while controlling for ICSEA, sex, staff retention rate, student enrollment, grade, and parental education. As can be seen in Table 9, the main effect of both time (indicating that support and safety changed in time) and group (indicating that there were differences in the level of support and safety between the PBL and control groups) were significant. These main effects are qualified by a significant interaction between time and group, which indicates that the groups differed in the evolution of support and safety in time. The PBL group showed a quadratic trend in time as seen in Figure 6 (F [1, 6668] = 3.80, p = .051; but not linear F [1, 6668] = 0.28, p = .597). The control group, in contrast, showed a linear trend in time (F [1, 6668] = 11.33, p = .001; but not quadratic, F [1, 6668] = 4.82, p = .028). We see a greater increase in support and safety in PBL schools compared to control groups.

Table 9. Results for support and safety

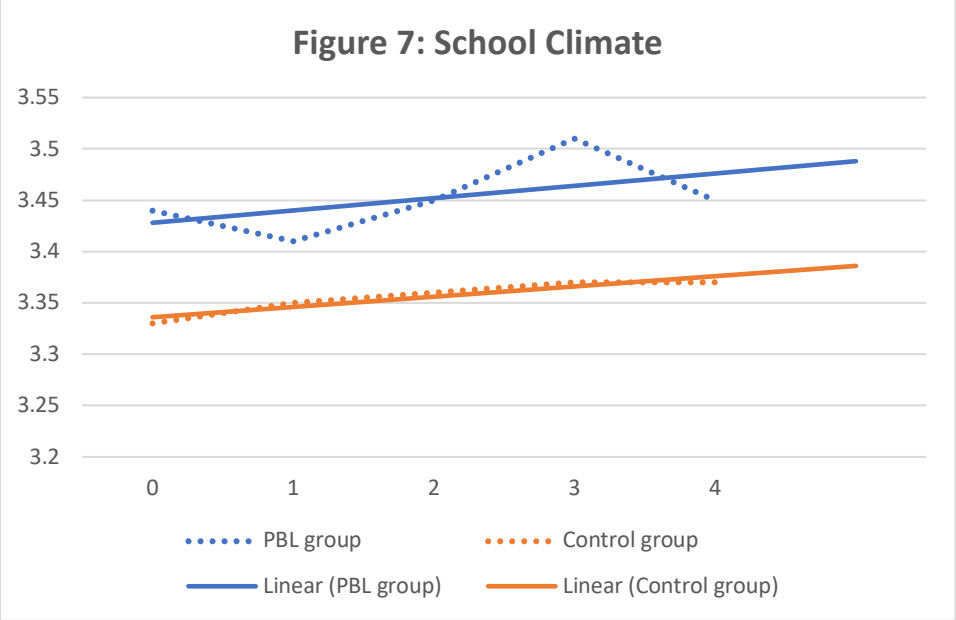
Predictor	Degrees of freedom	Mean Square	*F*	*p*
Intercept	1	5.29	3.75	.053
Group (Control = 0; PBL = 1)	1	29.07	20.63	<.001
Time (Year 0 to Year 4)	3.25	10.47	60.81	<.001
Group X Time	**3.25**	**7.70**	**44.71**	**<.001**
ICSEA	1	110.14	78.156	<.001
Sex	1	4.99	3.537	.060
Staff retention rate	1	43.24	30.685	<.001
Student enrollment	1	22.99	16.315	<.001
Grade	1	29.277	20.775	<.001
Parental Education	1	35.178	24.962	<.001
Error	6668	1.409		



School climate. To test whether students in schools that had PBL experienced an increase in school climate compared to students from the control group, a 5 (Time) X 2 (Group) mixed ANOVA was conducted with Time (year 0 to year 4) as the within-group factor and Group (control vs PBL) as the between-groups factor while controlling for ICSEA, sex, staff retention rate, student enrollment, grade, and parental education. Table 10 shows that there was only a main effect of Group, indicating differences between PBL and control groups, and a significant effect of time, indicating change in time. However, the interaction between time and group was non-significant. As can be seen in Figure 7, PBL schools showed on average higher school climate than the control group with school climate increasing in time.

Table 10. Results for school climate

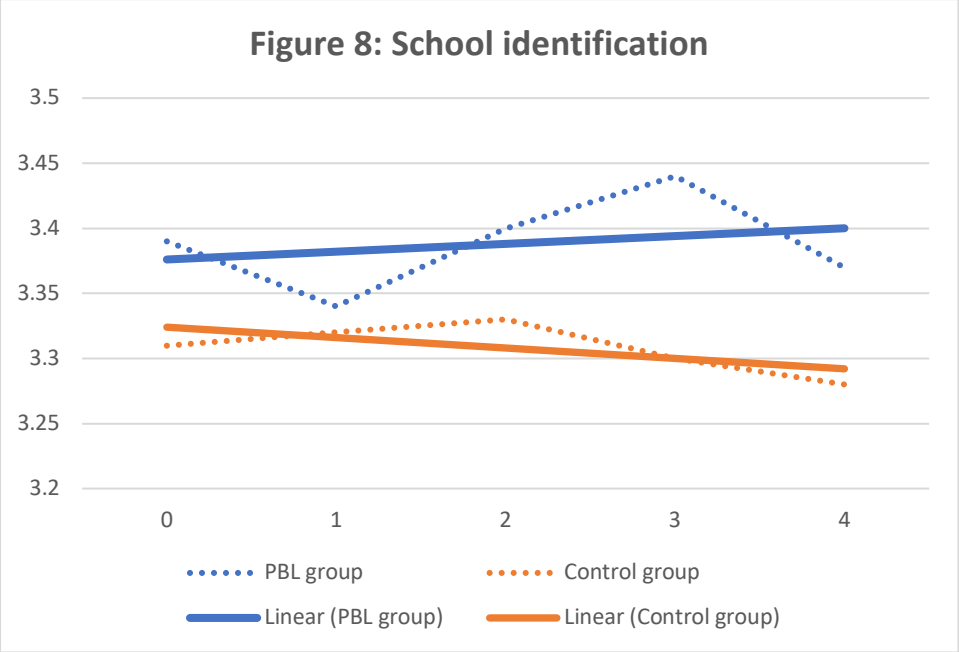
Predictor	Degrees of freedom	Mean Square	*F*	*p*
Intercept	1	52.66	90.91	<.001
Group (Control = 0; PBL = 1)	1	37.47	64.69	<.001
Time (Year 0 to Year 4)	3.97	1.43	3.07	.016
Group X Time	**3.98**	**.70**	**1.51**	**.198**
ICSEA	1	33.55	57.92	<.001
Sex	1	0.02	.03	.868
Staff retention rate	1	22.67	39.14	.000
Student enrollment	1	2.44	4.21	.040
Grade	1	29.277	20.775	<.001
Parental Education	1	35.178	24.962	<.001
Error	6668	1.409		



School identification. To test whether students in schools that had PBL experienced an increase in school identification compared to students from the control group, a 5 (Time) X 2 (Group) mixed ANOVA was conducted with Time (year 0 to year 4) as the within-group factor and Group (control vs PBL) as the between-groups factor while controlling for ICSEA, sex, staff retention rate, student enrollment, grade, and parental education. As can be seen in Table 11, the main effect of both time (indicating that school identification changed in time) and group (indicating that there were differences in the level of school identification between the PBL and control groups) were significant. These main effects are qualified by a significant interaction between time and group, which indicates that the groups differed in the evolution of school identification in time. The PBL group showed no change on time as seen in Figure 8 (linear, F [1, 6668] = 0.68, p = .140; quadratic F [1, 6668] = 0.06, p = .814) while the control group showed a linear decrease in school identification (F [1, 6668] = 4.25, p = .039; but not quadratic, F [1, 6668] = 1.05, p = .306). We see a decrease in school identification control groups while PBL schools remained stable.

Table 11. Results for school identification

Predictor	Degrees of freedom	Mean Square	*F*	*p*
Intercept	1	8.77	11.94	.001
Group (Control = 0; PBL = 1)	1	30.31	41.26	. <.001
Time (Year 0 to Year 4)	3.94	2.21	4.02	.003
Group X Time	**3.94**	**1.49**	**2.71**	**.029**
ICSEA	1	105.96	144.27	<.001
Sex	1	8.65	11.78	.001
Staff retention rate	1	41.95	57.11	<.001
Student enrollment	1	33.53	45.65	<.001
Grade	1	19.72	26.85	<.001
Parental Education	1	12.06	16.42	<.001
Error	6668	0.74		



The Effect of Fidelity of PBL

Depression. The effect of fidelity on depression was assessed using mixed modeling. We entered fidelity as a predictor of depression while controlling for year (or the time changes in depression), ICSEA, sex, staff retention rate, student enrollment, grade, and parental education. In addition, since every person has a different average level of depression, we allowed the intercept (mean level) and slope (the changes in time) of depression to vary across individuals.

First, a model without fidelity fit the data worse (-2RLL = 5199.27) than a model including fidelity (-2RLL = 3735.65; $\chi 2$ difference (1) = 1463.62, $p < .001$. This indicates that adding fidelity helped explain better the data. As can be seen in Table 12, individuals differed in their average level of depression (intercept) and the way depression changed in time (significant intercept and slope). Depression also increased as a function of Year (or time; significant effect of Year), but fidelity negatively and significantly predicted depression, indicating that as fidelity increased, depression decreased.

Table 12. Results for school identification

	Parameter estimate	Standard error	Wald *Z*/ *t*	*p*	95% confidence interval	
Predictor					**Lower bound**	**Upper bound**
Random effects			**Walds Z**			
Variance of the intercept	0.18	0.01	18.35	<.001	0.16	0.20
Slope	0.02	0.01	3.06	.002	0.01	.02
Covariance of intercept and slope	0.01	0.01	0.80	.422	-0.01	0.02
Fixed effects			***t*-test**			
Intercept	1.32	0.86	1.53	.127	-0.38	3.01
ICSEA	0.00	0.00	0.62	.532	0.00	0.00
Sex	0.15	0.03	5.22	<.001	0.09	0.21
Staff retention rate	-0.02	0.39	-0.06	.951	-0.78	0.73
Student enrollment	0.00	0.00	-0.31	.758	0.00	0.00
Grade	0.03	0.01	2.48	.013	0.01	0.05
Parental Education	0.00	0.03	0.09	.926	-0.05	0.06
Year	0.10	0.02	4.96	<.001	0.06	0.13
Fidelity of PBL	**-0.23**	**0.10**	**-2.38**	**.017**	**-0.43**	**-0.04**

Generalized anxiety. The effect of fidelity on generalized anxiety was assessed using mixed modeling. We entered fidelity as a predictor of anxiety while controlling for year (or the time changes in generalized anxiety), ICSEA, sex, staff retention rate, student enrollment, grade, and parental education. In addition, every person may differ in their average level of generalized anxiety, thus we allowed the intercept (mean level) and slope (the changes in time) of generalized anxiety to vary across individuals.

Results show that a model without fidelity fit the data worse (-2RLL = 4869.22) than a model including fidelity (-2RLL = 3456.93, $\chi 2$ difference (1) = 1412.29, $p < .001$. This indicates that adding fidelity helped explain better the data. As can be seen in Table 13, individuals differed in their average level of generalized anxiety (intercept) but not the way anxiety changed in time (slope). Generalized anxiety also increased as a function of Year (or time; significant effect of Year), but importantly, fidelity negatively and significantly predicted anxiety, indicating that as fidelity increased, anxiety decreased.

Table 13. Results for generalized anxiety

	Parameter estimate	Standard error	Wald *Z*/ *t*	*p*	95% confidence interval	
Predictor					**Lower bound**	**Upper bound**
Random effects			**Walds Z**			
Intercept	0.29	0.12	2.36	.018	0.13	0.66
Slope	0.01	0.01	1.24	.214	0.00	0.07
Covariance of intercept and slope	-0.04	0.04	-1.17	.242	-0.12	0.03
Fixed effects			***t*-test**			
Intercept	1.44	0.81	1.78	.075	-0.15	3.02
ICSEA	0.00	0.00	-0.04	.966	0.00	0.00
Sex	0.30	0.03	10.74	<.001	0.24	0.35
Staff retention rate	-0.10	0.36	-0.29	.773	-0.80	0.60
Student enrollment	0.00	0.00	0.09	.932	0.00	0.00
Grade	0.01	0.03	0.44	.661	-0.04	0.07
Parental Education	0.04	0.01	3.34	.001	0.01	0.06
Year	0.07	0.02	4.06	<.001	0.04	0.11
Fidelity of PBL	**-0.23**	**0.09**	**-2.48**	**.013**	**-0.42**	**-0.05**

Positive affect. The effect of fidelity on positive affect was assessed using mixed modeling. We entered fidelity as a predictor of positive affect while controlling for year (or the time changes in positive affect), ICSEA, sex, staff retention rate, student enrollment, grade, and parental education. In addition, since every person may have a different average level of positive affect, we allowed the intercept (mean level) and slope (the changes in time) of positive affect to vary across individuals.

First, a model without fidelity fit the data worse (-2RLL = 9612.30) than a model including fidelity (-2RLL = 6679.15; $\chi 2$ difference (1) = 2933.15, $p < .001$. This indicates that adding fidelity helped explain better the data. However, as Table 14 shows, fidelity did not significantly predict positive affect (with individuals differing in terms of their average level of positive affect [intercept] and the way positive affect changed in time [significant intercept and slope].

Table 14. Results for positive affect

	Parameter estimate	Standard error	Wald *Z*/ *t*	*p*	95% confidence interval	
Predictor					**Lower bound**	**Upper bound**
Random effects			**Walds Z**			
Intercept	1.88	0.54	3.52	<.001	1.08	3.29
Slope	0.11	0.05	2.05	.040	0.04	0.28
Covariance of intercept and slope	-0.34	0.16	-2.12	.034	-0.66	-0.03
Fixed effects			***t*-test**			
Intercept	4.24	1.72	2.46	.014	0.86	7.61
ICSEA	-0.16	0.04	-4.17	<.001	-0.24	-0.09
Sex	0.00	0.00	-0.06	.955	0.00	0.00
Staff retention rate	-0.22	0.06	-3.82	<.001	-0.34	-0.11
Student enrollment	0.62	0.76	0.82	.413	-0.87	2.11
Grade	0.00	0.00	0.19	.846	0.00	0.00
Parental Education	0.01	0.06	0.10	.917	-0.11	0.12
Year	-0.08	0.02	-3.72	<.001	-0.13	-0.04
Fidelity of PBL	**0.23**	**0.20**	**1.15**	**.252**	**-0.16**	**0.62**

Behavioral Engagement. The effect of fidelity on behavioural engagement was assessed using mixed modeling. We entered fidelity as a predictor of depression after controlling for year (or the time changes in behavioural engagement), ICSEA, sex, staff retention rate, student enrollment, grade, and parental education. In addition, since every person has potentially a different average level of behavioural engagement, we allowed the intercept (mean level) and slope (the changes in time) of behavioural engagement to vary across individuals.

Results show that a model without fidelity fit the data worse (-2RLL = 9213.02) than a model including fidelity (-2RLL = 6560.73, $\chi 2$ difference (1) = 2652.29, $p < .001$). This indicates that adding fidelity helped explain better the data. Despite this, fidelity did not significantly predict behavioural engagement (with significant individual differences in the intercept and slope, and year predicting less behavioural engagement in time).

Table 15. Results for behavioural engagement

	Parameter estimate	Standard error	Wald *Z*/ *t*	*p*	95% confidence interval	
Predictor					Lower bound	Upper bound
Random effects			**Walds Z**			
Intercept	0.56	0.21	2.73	.006	0.27	1.16
Slope	0.04	0.02	2.13	.033	0.02	0.11
Covariance of intercept and slope	-0.11	0.06	-1.72	.086	-0.24	0.02
Fixed effects			***t*-test**			
Intercept	6.77	1.07	6.34	<.001	4.68	8.87
ICSEA	0.00	0.00	-2.46	.014	0.00	0.00
Sex	0.12	0.03	3.55	<.001	0.05	0.19
Staff retention rate	0.36	0.43	0.83	.404	-0.48	1.20
Student enrollment	0.00	0.00	-0.29	.775	0.00	0.00
Grade	-0.10	0.01	-9.20	<.001	-0.12	-0.08
Parental Education	0.00	0.03	-0.03	.972	-0.07	0.07
Year	-0.11	0.02	-4.96	<.001	-0.16	-0.07
Fidelity of PBL	**0.12**	**0.11**	**1.06**	**.291**	**-0.10**	**0.34**

Emotional Engagement. The effect of fidelity on emotional engagement was assessed using mixed modeling. We entered fidelity as a predictor of depression after controlling for year (or the time changes in emotional engagement), ICSEA, sex, staff retention rate, student enrollment, grade, and parental education. In addition, since every person has potentially a different average level of emotional engagement, we allowed the intercept (mean level) and slope (the changes in time) of emotional engagement to vary across individuals.

Results show that a model without fidelity fit the data worse (-2RLL = 10598.) than a model including fidelity (-2RLL = 7620.25, $\chi 2$ difference (1) = 2973.21, $p < .001$). This indicates that adding fidelity helped explain better the data. However, as Table 16 shows, fidelity did not significantly predict emotional engagement (with individuals differing in terms of their average level of positive affect [intercept] and the way positive affect changed in time [significant intercept and slope].

Table 16. Results for emotional engagement

	Parameter estimate	Standard error	Wald *Z*/ *t*	*p*	95% confidence interval	
Predictor					**Lower bound**	**Upper bound**
Random effects			**Walds Z**			
Intercept	1.84	0.33	5.58	<.001	1.30	2.62
Slope	0.16	0.03	4.78	<.001	0.10	0.24
Covariance of intercept and slope	-0.46	0.10	-4.54	<.001	-0.67	-0.26
Fixed effects			***t*-test**			
Intercept	2.64	1.30	2.031	.042	0.09	5.18
ICSEA	0.00	0.00	1.431	.152	0.00	0.00
Sex	0.00	0.04	-0.053	.958	-0.08	0.08
Staff retention rate	-0.27	0.52	-0.514	.608	-1.28	0.75
Student enrollment	0.00	0.00	-0.745	.456	0.00	0.00
Grade	-0.08	0.01	-6.354	<.001	-0.11	-0.06
Parental Education	0.03	0.04	0.810	.418	-0.05	0.11
Year	-0.19	0.03	-6.765	<.001	-0.25	-0.14
Fidelity of PBL	0.018	0.14	0.130	.897	-0.25	0.28

Support and Safety. The effect of fidelity on support and safety was assessed using mixed modeling. We entered fidelity as a predictor of on support and safety while controlling for year (or the time changes in support and safety), ICSEA, sex, staff retention rate, student enrollment, grade, and parental education. In addition, every person may differ in their average level of support and safety, thus we allowed the intercept (mean level) and slope (the changes in time) of support and safety to vary across individuals.

Results show that a model without fidelity fit the data worse (-2RLL = 10266.70) than a model including fidelity (-2RLL = 7365.81, $\chi 2$ difference (1) = 2900, $p < .001$. This indicates that adding fidelity helped explain better the data. As can be seen in Table 17, individuals differed in their average level of support and safety (intercept) and the way it changed in time (slope). In addition, support and safety decreased as a function of Year (or time; significant effect of Year), but importantly, fidelity positively and significantly predicted support and safety, indicating that as fidelity increased, so did support and safety.

Table 17. Results for support and safety

	Parameter estimate	Standard error	Wald *Z*/ *t*	*p*	95% confidence interval	
Predictor					**Lower bound**	**Upper bound**
Random effects			**Walds Z**			
Intercept	1.47	0.29	5.00	<.001	0.99	2.17
Slope	0.11	0.03	3.92	<.001	0.07	0.19
Covariance of intercept and slope	-0.34	0.09	-3.76	<.001	-0.52	-0.16
Fixed effects			***t*-test**			
Intercept	3.21	1.25	2.58	.010	0.77	5.65
ICSEA	0.00	0.00	0.80	.424	0.00	0.00
Sex	-0.04	0.04	-1.10	.271	-0.12	0.03
Staff retention rate	-0.18	0.50	-0.36	.717	-1.15	0.79
Student enrollment	0.00	0.00	-1.29	.198	0.00	0.00
Grade	-0.07	0.01	-5.52	<.001	-0.09	-0.04
Parental Education	0.09	0.04	2.25	.025	0.01	0.17
Year	-0.16	0.03	-6.24	<.001	-0.22	-0.11
Fidelity of PBL	**0.27**	**0.13**	**2.09**	**.036**	**0.02**	**0.52**

School climate. The effect of fidelity on school climate was assessed using mixed modeling. We entered fidelity as a predictor of on school climate while controlling for year (or the time changes in support and safety), ICSEA, sex, staff retention rate, student enrollment, grade, and parental education. In addition, every person may differ in their average level of school climate, thus we allowed the intercept (mean level) and slope (the changes in time) of school climate to vary across individuals.

Results show that a model without fidelity fit the data worse (-2RLL = 7474.16) than a model including fidelity (-2RLL = 5784.52, $\chi 2$ difference (1) = 1689.64, $p < .001$. This indicates that adding fidelity helped explain better the data. As can be seen in Table 18, individuals differed in their average school climate (intercept) but not in how they changed in time (slope). In addition, school climate decreased as a function of Year (or time; significant effect of Year), but importantly, fidelity positively and significantly predicted school climate, such that as fidelity increased, so did school climate.

Table 18. Results for school climate

	Parameter estimate	Standard error	Wald *Z*/ *t*	*p*	95% confidence interval	
Predictor					**Lower bound**	**Upper bound**
Random effects			**Walds Z**			
Intercept	0.67	0.26	2.61	.009	0.31	1.42
Slope	0.05	0.03	1.91	.057	0.02	0.14
Covariance of intercept and slope	-0.11	0.08	-1.37	.170	-0.27	0.05
Fixed effects			***t*-test**			
Intercept	3.23	0.94	3.44	.001	1.39	5.08
ICSEA	0.00	0.00	0.93	.351	0.00	0.00
Sex	-0.02	0.03	-0.60	.546	-0.08	0.04
Staff retention rate	0.27	0.37	0.72	.470	-0.46	1.00
Student enrollment	0.00	0.00	-2.55	.011	0.00	0.00
Grade	0.06	0.03	2.08	.038	0.00	0.12
Parental Education	-0.09	0.01	-9.37	<.001	-0.11	-0.07
Year	-0.16	0.02	-8.05	<.001	-0.20	-0.12
Fidelity of PBL	**0.27**	**0.10**	**2.81**	**.005**	**0.08**	**0.46**

School identification. The effect of fidelity on school identification was assessed using mixed modeling. We entered fidelity as a predictor of on school identification while controlling for year (or the time changes in school identification), ICSEA, sex, staff retention rate, student enrollment, grade, and parental education. In addition, every person may differ in their average level of school identification, thus we allowed the intercept (mean level) and slope (the changes in time) of support and safety to vary across individuals.

Results show that a model without fidelity fit the data worse (-2RLL = 10318.68) than a model including fidelity (-2RLL = 7474.16, $\chi 2$ difference (1) = 2844.52, $p < .001$. This indicates that adding fidelity helped explain better the data. As can be seen in Table 19, individuals differed in their average school climate (intercept) but not in how they changed in terms of school identification in time (slope). In addition, school identification decreased as a function of Year (or time; significant effect of Year), but importantly, fidelity positively and significantly predicted school identification, indicating that as fidelity increased, so did school identification.

Table 19. Results for school identification

	Parameter estimate	Standard error	Wald *Z*/ *t*	*p*	95% confidence interval	
Predictor					**Lower bound**	**Upper bound**
Random effects			**Walds Z**			
Intercept	0.67	0.26	2.61	.009	0.31	1.42
Slope	0.05	0.03	1.91	.057	0.02	0.14
Covariance of intercept and slope	-0.11	0.08	-1.37	.170	-0.27	0.05
Fixed effects			***t*-test**			
Intercept	3.51	1.27	2.76	.006	1.02	6.00
ICSEA	0.00	0.00	1.03	.305	0.00	0.00
Sex	0.02	0.04	0.42	.678	-0.06	0.10
Staff retention rate	-0.70	0.51	-1.37	.170	-1.70	0.30
Student enrollment	0.00	0.00	-0.76	.447	0.00	0.00
Grade	-0.10	0.01	-7.93	<.001	-0.12	-0.07
Parental Education	0.08	0.04	2.03	.042	0.00	0.16
Year	-0.21	0.03	-7.88	<.001	-0.26	-0.15
Fidelity of PBL	**0.31**	**0.13**	**2.43**	**.015**	**0.06**	**0.57**