Introduction
In this case study, we summarise the impact measurement approach of Choo Choo Train, a programme developed by the Singapore Children’s Society (Children’s Society) for preschool children. What stands out about Choo Choo Train’s approach is that the team spent considerable effort in-house to research and develop an approach to short-term outcome measurement that even includes a control group as a part of the data collection.

Randomised Controlled Trials (comparing randomly selected programme participants to similar beneficiaries who did not participate in the programme) are often considered as the gold standard for impact measurement, but many charities feel that they lack the capacity or resources to carry out this type of study. This example shows how a charity can adopt some of the approaches associated with Randomised Controlled Trials to develop an evaluation that is relatively practical and affordable but may provide more compelling data than more basic approaches such as simple stakeholder surveys.

About Singapore Children's Society (Children’s Society)
Founded in 1952, Children’s Society was established during the post-war years of Singapore to provide shelter, comfort and relief to the children. Today, Children’s Society operates 11 service centres island wide, offering services in the four categories of: Vulnerable Children Services, Children and Youth Services, Family Services, and Research and Advocacy. The organisation’s mission is “to bring relief and happiness to children in need”.

About Choo Choo Train (CCT)
CCT is a structured community outreach programme aimed at instilling in five and six year old children the positive moral values of honesty, respect, compassion and responsibility. The programme sends CCT staff to partner pre-schools to deliver workshops using a combination of storytelling, role-plays and games. The programme structure was formalised in 2016 to respond to what Children’s Society identified as a changing cultural and societal context for young children’s moral education. As of 2017, 260 children have participated in the programme. In future CCT hopes to increase the range of values taught, and to train preschool teachers themselves to lead the programme.
How does Choo Choo Train measure its impact?
The diagram below shows a summary of the key steps in CCT’s impact measurement process:

**STEP ONE - Establishing Impact Goals: The team conducted thorough research, drawing on international frameworks as well as locally relevant research**

As a starting point, the programme team used the United Nations Human Rights Declaration to identify moral values that could be the focus of the programme. From this basis, they drew on their local research and experience to select four specific values (honesty, respect, compassion and responsibility) that they found to be most relevant in the Singapore context. One of the main sources that they used was a literature review conducted by the team on bullying in local schools. From their regular ground work with children and studying social trends, the team identified the importance teaching such values as an upstream measure to pre-empt social issues such as bullying. The four values were also selected to complement the pre-school curriculum, which has its own elements around learning pro social behaviours and cultivating positive relationships with peers.

Children’s Society has an in-house research team that supports frontline staff to conduct research and programme evaluation. For CCT, after selecting the four values, the research team worked with centre workers and a group of interns to establish a robust and relevant framework for the programme.

**STEP TWO - Developing the impact logic: the team used a logic model framework**

In 2016, the Children’s Society research team conducted logic model/ theory of change workshops for all programme staff, so that staff could have a clear way to set out the main target outcomes for their work. The CCT team laid out their programme logic in a table specifying “inputs”, “outputs”, “short-term outcomes” and “mid-term outcomes”, with arrows to show the links between each element (see below).
### CCT’s logic model framework

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Outputs</th>
<th>Short Term (1 month)</th>
<th>Mid Term (3 months)</th>
<th>Long Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. CCT team</td>
<td>-8 story-telling sessions for 4 different values.</td>
<td>-At least 85% of the children will be able to recall the values.</td>
<td>-Children display the values in their classroom interactions will want to apply the methods to make other’s feel better.</td>
<td>-Children will be equipped with a moral compass to guide their life.</td>
</tr>
<tr>
<td>2. Value based stories</td>
<td>-Five key discussion questions based on stories.</td>
<td>-At least 80% of children will participate in the discussion.</td>
<td>-Children are able to apply the values in hypothetical scenarios.</td>
<td>-Children will have social awareness and able to build better social relationships.</td>
</tr>
<tr>
<td>3. Time</td>
<td>-Application of value through role play, games and worksheets.</td>
<td>-At least 80% of children complete the activities.</td>
<td>-Children are able to share methods on application of values.</td>
<td></td>
</tr>
<tr>
<td>4. Materials</td>
<td>-Teachers will reinforce the values and teachings in their day to day interaction with the children</td>
<td>-At least one preschool teacher to sit in every session.</td>
<td>-At least 80% of the preschool teachers affirm child on behaviour based on value.</td>
<td></td>
</tr>
<tr>
<td>5. Research</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Preschool</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teachers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Children</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Parents (?)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Assumptions**
- Children will be engaged in the sessions.
- Ample opportunities for the children to apply the values.

**External factors**
- Parents model the values at home.

---

Rising trend of bullying in Singapore
STEP THREE - Selecting tools and metrics: The team identified existing measurement tools which they adapted and tested to suit the context

Staff, volunteers and interns did intensive research to understand existing outcome measures for similar programmes. They found that most of the indicators used were not relevant as they measured different aspects of the values, instead of what CCT intended to emphasise. Nevertheless, they found three existing measurement tools which they were able to adapt: I) “moral puppet interview”\(^1\), II) teacher rating scale\(^2\) and III) scenario based questions\(^3\).

Since the CCT programme was targeted at five and six year olds, the team needed to ensure the data collection method was engaging and clearly understood by the children. It was also important that it should align with the content and context of the programme. CCT contacted the original creator of “moral puppet interview” to seek advice on implementation. Meanwhile, as they could not find scenario based questions that were specific to the testing values for the programme, they crafted their own based on the stories/curriculum developed by programme staff. This helped to ensure that the scenarios were not too remote from what was taught to the children.

Before rolling out the first run of the CCT, the team did a pilot test on the receptiveness of those data collection methods, involving preschool children from another Children’s Society programme. Initially, the “moral puppet interview” was recorded on video as a tool for administering pre and posts tests. However, the screened video of “moral puppet interview” was found to be not engaging enough and was changed to real life hand puppets. Meanwhile, the original images to support the scenario based questions were found to have too many distractions, and so the CCT staff created new standardised images.

**Moral Puppet Interview:**

Programme staff ask children which puppet is more similar to themselves, then fill in the form accordingly.

---

\(^1\) Kochanska et al. (2010)

\(^2\) Adapted from Weir & Duveen’s (1981) Prosocial Behavior Questionnaire (Teacher-rating), and two social competence scales used in Fast Track (Teacher-rating)

\(^3\) Based the format on the Wally’s Test (Webster-Stratton, 1998) used by the Incredible Years programme.
STEP FOUR - Collecting data: The team collected data pre- and post- intervention from a small sample of participants as well as a small sample of non-participants acting as a control group

The team collected data from two groups of children: an intervention group (who participated in the programme) and a control group (similar children who did not participate in the programme). The intervention group was a random sample of 17 of the 233 children who took part in the programme. The control group was a random sample of 19 children from a nearby preschool that was not yet participating in the programme.

The reason for taking this two group approach was to identify a “counterfactual”: in other words, to show not only what change had taken place for the children participating in the programme, but also give an indication of whether that change would have taken place anyway even for similar children who did not participate.

Data was collected from each child two weeks before the start of the programme and once again two weeks after the end of the eight-week programme. The data consisted of the two types of children’s self-report described in the previous section, as well as a teacher observation form.

STEP FIVE - Analysing Data: The team triangulated the pre- post- data with informal observations from programme staff but they were cautious about the findings due to the small sample size

To analyse and interpret the data, the team triangulated findings from three sources:

- The quantitative data from the children (moral puppet interviews and scenario based interviews).
- The teacher observations provided through the structured observation form.
- More informal observations from the programme staff who had interacted directly with the children over the course of the programme.

For the children’s data, results indicated that the intervention group had slightly higher levels of improvement than the control group for all four moral values. While the teachers’ observations...
August 2017

indicated that the intervention group showed slightly higher levels of improvement than the control group for two of the four values, responsibility and compassion.

The team was able to draw the tentative conclusion that the programme had succeeded in enabling children to learn and better apply the values in their daily interactions. However, the small sample size meant that they were cautious about their findings.

**STEP SIX - Using the Findings: Two main ways - internally to help iterate and improve the programme; and externally to engage stakeholders**

Internally, the CCT programme team on the ground continuously reflects on progress and lessons learned to iterate and improve their approach. On a day-to-day basis, they use their own observations at the end of each session. The impact measurement results from year one also fed into this cycle of continuous improvement: when the results were ready, the research and programme teams sat down together to discuss and learn from the findings.

One example of how the team refined the programme relates to the finding that the children did not seem to show a significant improvement for the value of honesty. The entire CCT team discussed this result before making changes to honesty sessions in the second run in 2017 – by changing the format of stories told. Another example relates to the feedback from teachers about the programme content and facilitation. Based on the teachers’ feedback, CCT updated the programme to teach simpler and clearer definitions of the different values that would be easier for the children to understand.

In general, at Children’s Society this process of data review and continuous improvement tends to take place at the frontline level rather than being explicitly driven by reporting requirements at Board or senior management level. It is part of a broader culture of accountability and learning that the organisation aims to nurture at all levels. One way that Children’s Society promotes this way of thinking is through its Programme Evaluation Workshop, an event that takes place every two years where staff from each centre within Children’s Society present an evaluation of one of their services. External experts are also invited to these events to provide their comments.

“As a charity we have to be responsible for the standard of our programmes, internally we also want to ensure that we provide quality.”

The team’s second main use for their impact data is for engaging with external stakeholders. Children’s Society staff have found over the years that different types of funder and partner demand different levels of data. In some cases, they may be asked simply to show evidence that the need for a programme exists, without having to prove the actual efficacy of the programme itself. But in other cases, it is important to be able to show compelling evidence of actual impact. The team note that impact data is particularly helpful when you want to scale up a programme and encourage others to adopt it. For CCT, the ultimate vision is for the programme to scale up and be adopted by schools themselves – for this reason, the impact data will be important in due course for convincing stakeholders such as school principals and relevant government agencies of the broader value of the programme.
What can we learn from Choo Choo Train’s impact measurement approach?

Main challenges

Time consuming
The team found that the process of designing and implementing the evaluation was extremely time consuming. The bulk of the time was spent on researching and developing the tools before the start of the programme; meanwhile the task of collecting data also took considerable time.

Not getting it right first time meant that some data was wasted
As in many cases, the evaluation tools did not work perfectly the first time round. Some questions were not understood by the students, or the children were not engaged enough to answer them properly. With the reliability of data being affected, this meant that some data could not be used in the analysis for year one. However, the team commented that it is important to see the evaluation as an iterative process where the tools will continue to be refined over time and it is unrealistic to expect the exercise to run perfectly smoothly first time round.

Small sample size and lack of randomisation: implications for rigour
Even though the intervention group displayed an improvement in pre-post test scores, the CCT team acknowledges that the sample size in 2016 was too small to claim statistically significant results between the control and intervention group. Also, they are aware that due to the different schools having varying demographics of students and teachers, their controlled trial should not be considered as truly “randomised”. They are currently looking for more preschools that are willing to participate as control group centres so as to strengthen the quality of data in the next round.

Finding Control Groups
It was challenging for the team to identify an appropriate group of children to act as the control group. To get the most robust results, the children in the control group should have similar characteristics and experiences to those in the intervention group so that the main difference between the two groups is the fact that one participated in the programme whilst the other did not. In this sense, the best comparison would have been to split each class into two groups where one received the programme and the other did not. However, this was not feasible for either practical or ethical reasons. Therefore, the team looked for a nearby school with similar aged children to act as the control group.

They found that some schools were not willing to act as the control group because the data collection would have created a logistical burden on teachers by taking up class time (even though the data collection itself was done by Children’s Society staff). One local school ultimately agreed to participate as control group on the understanding that their children would then actually receive the programme in the following year.

However, the CCT team noted during the course of the evaluation that even though the intervention and control group schools were nearby, there were some significant differences in the teaching and management styles – which partly limited their ability to draw conclusions from the data.
What worked well

**Having a logic model made the target outcomes clear from the start**
Due to the logic model, all the programme facilitators understood how each component led to the short-term and mid-term target outcomes. It made delivering the programme more streamlined, as staff knew what they were looking to achieve.

**Ground staff having a close partnership with the research team**
The team commented that mutual understanding and clear communication between the research team and practitioners is crucial for the success of the evaluation. They noted that the research team needs to understand that realistically, the data collected might not be perfect; and practitioners need to want to see their programme work well. It was also important for the research team to use clear language and not to speak in technical jargon that would be confusing for the programme team. With this strong partnership in place, frontline programme staff found that the evaluation helped provide additional focus and motivation for the team.

“The structure of impact evaluation gave us motivation to improve.”
(Programme team member)

**Using a strong team of interns and volunteers**
The evaluation team was supported by four interns and volunteers who worked part-time on the project for several months. Children’s Society’s research team commented that it would not have been possible to take such a thorough approach without the additional manpower provided by interns and volunteers.

**Building on existing evaluation experience**
As Children’s Society conducts evaluations on an ongoing basis with many of its programmes, the research team had a bank of knowledge and experience to help inform their approach with CCT. In particular, there was one similar programme which had already undergone many iterations of the evaluation tool and offered many lessons learned to support CCT’s evaluation design.

“The reality is that the process of programme improvement will never end. It takes time, effort, and most of all patience.”

Caveat: This case study is provided as an example of how one organisation has approached impact measurement in one particular context. Please note that readers should not interpret it as “best practice” as there are many valid approaches in any given context. The information has been kindly shared by the organisation to help others in the sector learn more about what worked well and the challenges faced with the different approaches to impact measurement.